THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à MINES ParisTech

Safety benefit assessment, vehicle trial safety and crash analysis of automated driving : a Systems Theoretic approach

Evaluation des gains de sécurité, sécurisation des essais et analyse des accidents du véhicule autonome : une approche systémique

Ecole doctorale n°432

SCIENCES ET METIERS DE L'INGENIEUR

Spécialité : SCIENCES ET GENIE DES ACTIVTES A RISQUES

Stephanie **ALVAREZ GOMEZ**

Dirigée par **Franck GUARNIERI**

Encadrée par **Yves PAGE**

COMPOSITION DU JURY:

M. Paul SALMON
University of the Sunshine Coast, Rapporteur

M. Enrico ZIO
Politecnico di Milano, Rapporteur

Mme. Nancy LEVESON
Massachusetts Institute of Technology, Examinateur

M. Pierre VAN ELSLANDE
IFSTTAR, Examinateur

M. Franck GUARNIERI
MINES ParisTech, Examinateur

M. Yves Page
Renault, membre invité

MINES ParisTech | PSL RESEARCH UNIVERSITY PARIS

Table of Contents

List of Figures

List of Tables

Résumé chapitre 1: Introduction

Ce chapitre d'introduction de la thèse présente, le contexte, la problématique et les questions de recherches en découlant.

Trois bouleversements technologiques sont en cours dans la mobilité, passage à une propulsion électrique, développement de la connectivité du véhicule et enfin automatisation de la conduite, avec de nombreux bénéfices attendus. En particulier, l'amélioration de la sécurité routière est une des justifications avancées au développement du véhicule autonome. Aujourd'hui, plus de 1,2 millions de personnes décèdent tous les ans dans un accident de la route dans le monde et la sécurité routière est une préoccupation majeure pour de nombreux gouvernements.

Le propos de la thèse est d'étudier l'influence du véhicule autonome sur la sécurité routière au travers des trois questions de recherche identifiées :

i. Les systèmes de conduite automatisée vont-ils améliorer la sécurité routière ?
ii. Comment sécuriser les expérimentations du véhicule autonome ?
iii. Comment analyser les accidents de la route impliquant des systèmes de conduite automatisée ?

Un cadre conceptuel est nécessaire afin de répondre à ces questions et la littérature suggère que les méthodes existantes dans la sécurité routière sont dépassées et qu'un changement de paradigme vers la théorie des systèmes est nécessaire. En réponse, l'objectif de la thèse est d'aborder les trois questions de recherche en appliquant une approche basée sur la théorie de systèmes.

Dans ce chapitre, nous retrouvons également le plan d'approche de la thèse et la structure du manuscrit.

Chapter 1: Introduction

1.1 Problem statement

Currently, three disruptive innovations: electrification, connectivity and automation, are transforming the future of mobility. Electric vehicles are expected to contribute to a clean mobility by reducing petroleum consumption, increasing fuel efficiency, and decreasing greenhouse gas emissions[1]. Connected vehicles offer new services and applications that can bring potential benefits such as improving traffic flow, enhancing traffic cooperation, and improving the comfort and safety of road users. To do this, connected vehicles use wireless short-range communications between a vehicle and other vehicles (V2V), infrastructure (V2I), and others including pedestrians or the cloud (V2X). Lastly, automated vehicles are expected to improve road safety, gain comfort and time for the driver, provide mobility for everyone, improve fuel consumption, etc. by taking over part or all of the dynamic driving task.

On the other hand, around 1.2 million people are killed on the roads every year and between 20 and 50 million suffer injuries (World Health Organization 2015). Road crashes are the leading cause of death among people aged 15 to 29 years. Further, road traffic injuries are the ninth leading cause of death across all age groups and are predicted to become the seventh leading cause by 2030 (World Health Organization 2015). As a response, countries all around the world must establish and review road safety strategies and measures to reduce the current unacceptable levels of road trauma.

All vehicle technologies, including the three aforementioned innovations have the potential to influence road safety by affecting exposure, crash risk, and crash consequences (Risto Kulmala 2010). For example, (Minelli, Izadpanah, and Razavi 2015) used micro-simulation models to demonstrate that connected vehicles have an effect on exposure by increasing travel times.

[1] The greenhouse gas emissions of an electric vehicle also depend on the source of fuel and technology employed for generating energy.

Accordingly, automakers have the responsibility to address the safety implication of these innovations throughout the entire development and operation process of their vehicles.

The French-based automaker Renault (the industrial partner of the thesis) identified three challenges regarding the implication of automated driving systems on safety during several phases of the process. These challenges led to the definition of the research questions that motivated the research conducted in this thesis.

1. The safety benefit assessment of automated driving systems during the validation phase:

Automated driving is expected to improve road safety by eliminating human driver error which has been attributed as the main cause of crashes (Treat 1977); however, automation may also introduce new hazards such as the inadequate operation of automation's perception system and unsafe driver behavior during driving transition phases (e.g. transition from automated driving mode to manual driving mode). Consequently, automakers must assess the safety benefit of automated driving systems in terms of crash avoidance and injury mitigation, to confirm their safety impact during the validation phase. Further, the validation phase also comprises providing evidence on the safety benefit of automated driving systems to policy-makers, customers, opinion leaders, etc. to demonstrate that their expected safety benefits are real.

The challenge related to the safety benefit assessment generated the first research question of this thesis:

i. **Will automated driving improve road safety?**

2. The safety of vehicle trials during the deployment phase of automated driving systems:

The deployment of automated driving requires testing the systems through multiple means such as component-level testing, testing on driving simulators, on closed-track and on open roads, in order to demonstrate the safety of automated driving systems. The vehicle trials involving real-driving on closed-tracks and open roads are particularly hazardous due to the risk of crashes, injury to road users and property damage.

The need to consider safety of automated driving trials led to the definition of the second research question:

ii. How to ensure the safety of automated driving trials?

3. The analysis of crashes involving automated driving systems during the retrospective evaluation phase

The assumption that automated driving systems reduce the number of crashes implies that the nature of the remaining crashes may be different. Therefore, automakers may need new crash analysis methods suitable for automated driving for the retrospective evaluation of automated driving systems. In fact, the existing crash analysis methods are focused on the human driver as the last regulator of the system; removing the driver from the control loop and giving the vehicle the possibility to be the last regulator, brings different risks which may need to be analyzed differently.

The challenge associated to the analysis of crashes involving new types of interactions and risks introduced by vehicle automation, led to the definition of the third research question:

iii. How to analyze road crashes involving automated driving?

In order to address these research questions (and other questions related to automated driving and road safety), a conceptual framework is needed; however, the changes introduced by automation to the interactions between the main components of the road transport system (i.e. the driver, the vehicle and the environment) challenge the capability of the conceptual frameworks currently used in road safety which have a main focus on the human driver, to comprehensively understand the new system. It was assumed that the existing road safety methods are not capable of comprehensively assisting the analysis of road transport systems with vehicle automation, and that a new conceptual framework is needed.

Several studies have suggested that road safety should shift to a systems theory paradigm in order to support the analysis and understanding of the increasingly complex road transport sociotechnical system (Larsson, Dekker, and Tingvall 2010; Salmon, McClure, and Stanton 2012; Salmon and Lenné 2015). While first applications of systems theory to road safety have already been demonstrated on today's road transport system (K. L. Young and Salmon 2015; Scott-

Parker, Goode, and Salmon 2015; Salmon, Read, and Stevens 2016), the application to road transport systems involving vehicle automation has not yet been examined[2]. Consequently, it seems clear that the use of systems theory as a conceptual framework to address the three research questions should be investigated.

1.2 Research aims

This thesis aims to examine the safety benefit, trial safety and accident analysis of automated driving by applying a systems theoretic approach suitable for vehicle automation.

- Identify the systems theoretic conceptual framework by analyzing the literature.
- Adapt and extend the conceptual framework to meet the needs of automated driving and road safety.
- Use the systems theoretic approach and its extensions to examine the safety benefit, trial safety and accident analysis of automated driving systems.

1.3 Research approach

The literature on systems theoretic approaches was reviewed in order to identify a model called Systems Theoretic Accident Model and Processes (STAMP) as the specific conceptual framework used in the thesis. To contribute to the safety benefit assessment of automated driving vehicles, the target population of a highway pilot system was estimated and an STPA (hazard analysis method based on STAMP) analysis on the highway pilot system was conducted to assist the evaluation of the system effectiveness. Subsequently, trial safety was examined by performing two STPA analyses: firstly on the vehicle trial process and secondly on an automated driving trial involving a highway pilot system. Finally, the process to analyze crashes involving automated driving systems was investigated by extending CAST (accident analysis method based on STAMP)

[2] There have been studies that apply systems theory on automated driving systems to evaluate safety at the vehicle level and the interactions with the human driver (Van Eikema Hommes 2012), but to my knowledge, there are no applications of systems theory to road safety and vehicle automation at the traffic system level.

into a method called CASCAD (Causal Analysis using STAMP for Connected and Automated Driving) which is adapted to vehicle automation.

1.4 Thesis structure

1.4.1 Chapter 1

The introductory chapter presents the problem statement, the research aims, the approach used in the thesis and the structure of the thesis.

1.4.2 Chapter 2

The literature review that provides the context and the conceptual framework for the thesis is described in the second chapter. The chapter is organized according to three main topics: vehicle automation, road safety and the systems theoretic approaches to safety. After the description of vehicle automation and road safety, the evidence from the literature indicating the suggestion to move towards a systems theoretic approach to road safety and an overview of the three most popular systems theoretic approaches are presented. Next, the Systems-Theoretic Accident Model and Processes (STAMP), STPA (hazard analysis method based on STAMP) and CAST (accident analysis method based on STAMP) are selected as the contextual framework to address the three research questions. Lastly, the model STAMP, and the two methods are described in detail.

1.4.3 Chapter 3

The safety benefit assessment of automated driving systems is examined in this chapter using the case study of a highway pilot system. To this end, a contribution to the safety benefit assessment of automated driving systems is provided by estimating the target population of the highway pilot system and by elaborating of questions to facilitate the evaluation of the effects of the highway pilot system on road safety. These questions are elaborated based on the outputs of an STPA analysis on the highway pilot system.

1.4.4 Chapter 4

The safety of automated driving trials is addressed in chapter four. A first STPA analysis is conducted on the vehicle trial process at a macroscopic level (the analysis considers high-level controllers such as the government and the vehicle company management). Next, a second STPA analysis is conducted on a vehicle trial operation involving a highway pilot system at the microscopic level (only the actors in the low levels of the operating process are considered). The results of the two analyses are organized to create a framework to ensure the safety of automated driving trials.

1.4.5 Chapter 5

This chapter examines the analysis of crashes involving automated driving by introducing CASCAD (Causal Analysis using STAMP for Connected and Automated Driving), an extension of CAST for this type of crashes. Two current accident analysis techniques from road safety are first described to identify elements specific to road safety which can be transferred to the analysis of crashes involving vehicle automation. Next, STAMP concepts are used to develop guidance elements that facilitate the application of CAST on road crashes involving automated driving. The two types of elements are integrated into CAST in order to create CASCAD. Lastly, the application of CASCAD is illustrated[3] with a real-world crash.

1.4.6 Chapter 6

Chapter six provides a discussion of the research conducted in this thesis. Initially, a summary of the findings in chapters 3-5 is presented. The main contributions of the research as a whole are provided, namely the implications of STAMP-based methods for the three research questions, the representation of the road transport system as a control structure, the larger scope of the analysis and identified causal factors, and the modifications developed in the thesis to apply STPA and CAST. Finally, the methodological considerations of the thesis are discussed.

[3] The CASCAD analysis on the real-world crash is intended to illustrate the CASCAD process, not to conduct a complete analysis of the crash.

1.4.7 Chapter 7

The last chapter presents the overall conclusions of the thesis relative to the separate findings on the three research questions and the contribution of the research as a whole. Further, future work is proposed including the progression of the research conducted in chapters 3-5 regarding the three research questions, perspectives on the use of a STAMP-based approach for new automated driving systems and new safety applications, and suggestions to encourage the adoption of this approach by the road safety community.

Résumé chapitre 2: Véhicule autonome, Sécurité Routière et Approches fondées sur la théorie des systèmes

Le second chapitre présente la revue de la littérature axée autour de trois sujets, le véhicule autonome, la sécurité routière et les approches fondées sur la théorie des systèmes, donnant le contexte et le cadre conceptuel de la thèse.

Le véhicule autonome est abordé à travers la taxonomie des différents niveaux d'automatisation, les principales motivations, les stratégies et les divers challenges liés à son développement.

Concernant la sécurité routière, deux définitions sont fournies : la sécurité routière en tant qu'insécurité sur les routes et en tant que système. Ensuite, les différents points de vue de la sécurité routière au fil du temps sont décrits. La référence en termes de politique de sécurité routière « Safe system approach » est exposée.

Trois approches basées sur la théorie des systèmes sont ensuite décrites: le Risk management Framework, STAMP et FRAM. La dernière partie est dédiée aux descriptions plus détaillées de STAMP, STPA et CAST, les modèles et les méthodes sélectionnés comme le cadre conceptuel utilisé pour l'évaluation des gains de sécurité, la sécurisation des expérimentations et l'analyse des accidents des systèmes de conduites autonomes.

Chapter 2: Vehicle automation, Road Safety and Systems Theoretic approaches

2.1 Chapter overview

The second chapter presents a review of the literature on three tropics: vehicle automation, road safety and the systems theoretic approaches to safety, which provides the context and the conceptual framework for the thesis (as illustrated in figure 1). The subsection on vehicle automation includes the automated driving taxonomy, and the main motivations, paths and challenges of vehicle automation. The subsection on road safety provides two definitions of road safety (road safety as the lack of safety and road safety as a system), the road safety perspectives over time, and introduces the Safe System approach[4]. The subsection on systems theoretic approaches first examines the studies that have recommended adopting a systems theory approach to road safety.

Next, three approaches based on systems theory are described: the Risk Management Framework, STAMP and FRAM. The last subsection is dedicated to the detailed description of STAMP, STPA and CAST, which constitute the model and methods selected as the conceptual framework to address the safety benefit, trial safety and accident analysis of automated driving systems.

[4] The Safe System approach is the vision behind road safety strategies introduced by countries like Sweden; it does not imply systems theory.

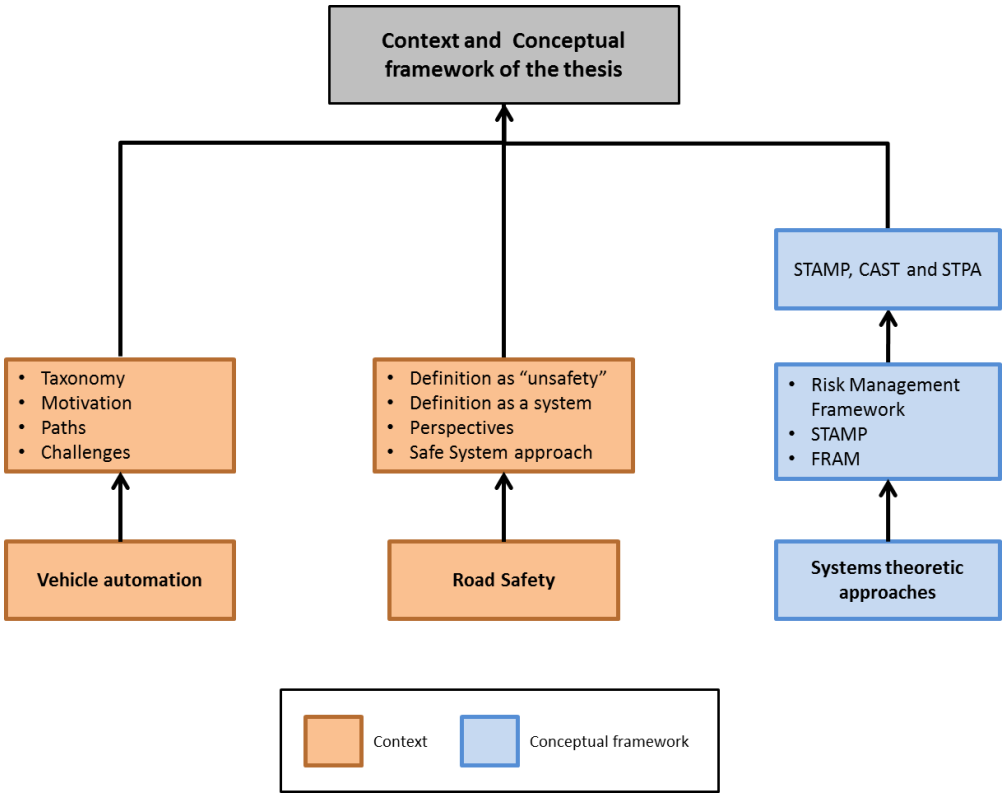


Figure 1 – Context and conceptual framework of the thesis

2.2 Vehicle automation

Section 2.2 describes the definitions and taxonomies for vehicle automation, the expected benefits that motivate vehicle automation advocates, and the paths and challenges for the development of this type of technology.

2.2.1 Vehicle automation definition and taxonomy

Vehicle automation involves a range of several levels in which driving tasks previously performed by the human driver are progressively delegated to driving automation systems. Therefore, vehicle automation is described though taxonomies of the range of levels of driving automation. There are three main taxonomies for vehicle automation: a taxonomy established by the German Federal Highway Research Institute (Gasser and Westhoff 2012), a taxonomy defined by the National Highway Traffic Safety Administration (NHTSA) in the US (NHTSA 2013), and finally the taxonomy provided by the Society of Automotive Engineers (SAE) (SAE International 2012, 2016). Although the three taxonomies of vehicle automation levels have

many similarities (see the two last columns of table 1 for the correspondence between the levels in the three taxonomies), the SAE is the most widely used.

The SAE taxonomy classifies vehicle automation systems that perform a part or all of the dynamic driving task (DDT) on a sustained basis, in 6 levels of driving automation which range from no driving automation (level 0) to full driving automation (level 5). The levels refer to the driving automation features that are engaged during the operation of vehicle; consequently, a given vehicle can be equipped with a driving automation system capable of engaging multiple features. Moreover, information systems, active safety systems e.g. automated emergency braking, and certain assistance systems such as lane keeping assistance, which do not perform part of the DDT on a sustained basis, but provide a momentary intervention, are excluded from the classification.

Aspects considered in the taxonomy

The taxonomy classifies the levels of automation according to the roles of the human driver and the driving automation system, relative to three aspects:

1. The Dynamic Driving Task (DDT).
2. The Dynamic Driving (DDT) task fallback.
3. The Operational Design Domain (ODD).

Dynamic Driving Task (DDT)

The Dynamic Driving Task (DDT) refers to the three hierarchical levels of the driving task established by (Michon 1985) strategic, maneuvering and control (illustrated in figure 2) to define the scope of the DDT which includes the maneuvering and control levels, but excludes the strategic level.

At the strategic level, activities are related to planning and executing a trip from origin to destination. The need for processing information only occurs occasionally, with intervals ranging from a few minutes to hours. The decisions made at this level provide inputs to the next level. The maneuvering or tactical level refers to tasks dealing with the interaction with both the environment and other road users. Activity is required rather frequently, with intervals

from a few seconds to a few minutes. It refers to elements such as speed choice, lane choice and provides the input for the lowest task level. Finally, at the control or operational level, the motion of the vehicle is controlled in the longitudinal and lateral direction, and information has to be processed frequently, ranging from intermittent activities every few seconds to almost continuous control.

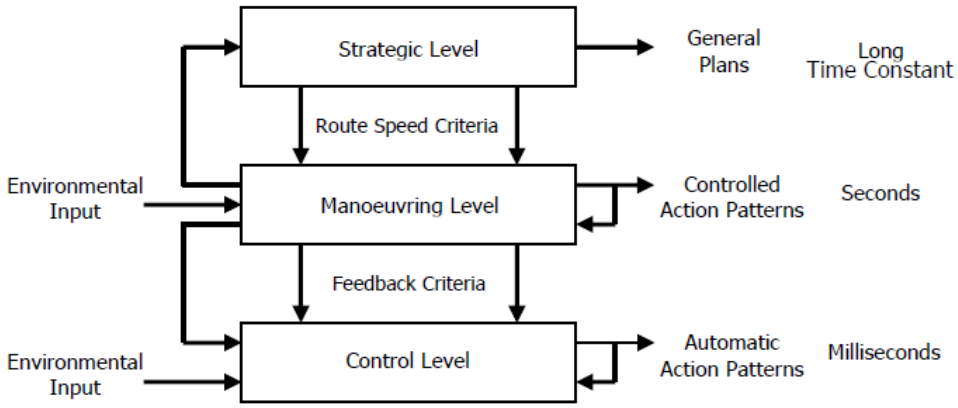


Figure 2 - The three levels of the driving task according to (Michon 1985)

Accordingly, the SAE defines the DDT defined as “all of the real-time functions required to operate a vehicle in on-road traffic, excluding the strategic functions such as trip scheduling and selection of destinations and waypoints (i.e., navigation or route planning), and including without limitation":

1. Lateral vehicle motion via steering (control level): it includes the detection of the vehicle positioning relative to lane boundaries and the application of steering to maintain the vehicle in an appropriate lateral positon.
2. Longitudinal vehicle motion via acceleration and deceleration (control level): It includes setting speed and the detection of a preceding vehicle (if any), and the application of acceleration or braking to maintain speed or an appropriate gap to the preceding vehicle.
3. Monitoring the driving environment via object and event detection, recognition, classification and response preparation (control and maneuvering levels).
4. Object and event response execution (control and maneuvering levels).

For simplification purposes subtasks (3) and (4) are grouped in the term Object and Event Detection and Response (OEDR).

5. Maneuver planning (maneuvering level); and
6. Enhancing conspicuity (easily seen or noticed) via lighting, signaling and gesturing, etc. (maneuvering level).

Figure 3 displays a graphical view of the driving task which includes functions at the strategic, maneuvering and control levels. The portions of the driving task covered by the dynamic driving task are illustrated inside the blue box.

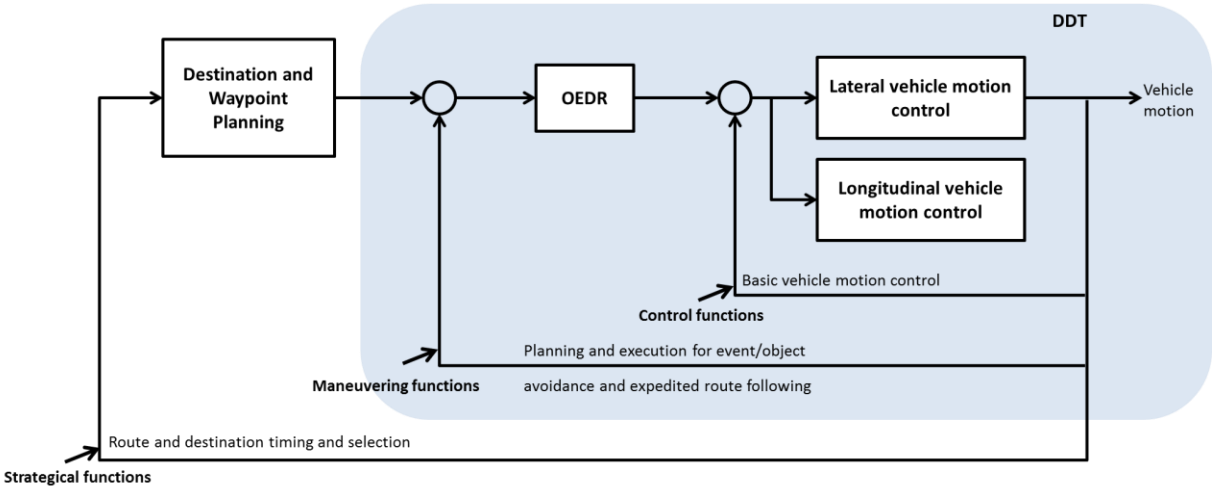


Figure 3 – Schematic view of driving showing DDT portion adapted from (SAE International 2016)

Dynamic Driving Task (DDT) Fallback

The dynamic driving task fallback is described as the response by the human driver or automation to either perform the dynamic driving task or achieve a minimal risk condition (i.e. condition to reduce the risk of a crash) after a relevant failure or malfunction in a driving automation system and/or other vehicle system, or upon the exit of the operational design domain.

An example of DDT fallback by performing the DDT is when the driver takes over the DDT after a vehicle sensor failure that prevents automation from continuing to perform the DDT. Further, an example of DDT fallback by achieving a minimal risk conditions is when automation removes

the vehicle form the active lane of traffic before coming to a stop as it reaches a highway exit (for an automated driving system that can only be operated in highways).

Operational Design Domain (ODD)

The Operational Design Domain (ODD) includes the specific conditions under which a driving automation system is designated to function. These conditions include geographic, roadway, environmental, traffic and speed limitations. For instance, an automated diving system may be designed to be exclusively operated on highways, within a speed range of 10-90 km/h, under heavy traffic, during daytime only and with no heavy rain.

Driving automation taxonomy:

Table 1 shows a summary of the SAE taxonomy consisting of six discrete and mutually exclusive levels (SAE International 2012, 2016). As seen in the table, the roles of the human driver and automation (i.e. ADS system) relative to the three aspects previously described: DDT, DDT fallback and ODD, are central to categorize the six levels.

At level 0, the driver is expected to perform the entire DDT and DDT fallback. At levels 1 and 2, the Automated Driving System (ADS) performs either the longitudinal and/or lateral vehicle motion control subtask of the DDT, while the driver performs the remaining vehicle motion control task (for level 1), the OEDR task of the DDT, and the DDT fallback. Moreover, the ODD of the system is limited.

Level 3 establishes a rupture as the ADS performs the entire DDT. However the driver is expected to perform the DDT fallback and the ADS has a limited ODD. The main difference between level 3 and level 4 is the capability of the ADS to perform DDT fallback.

In level 5, the ADS performs the entire DDT, DDT fallback and does not have a prescribed ODD. Finally, the last two columns of the table 1 display a comparison to the classification levels defined in the other two taxonomies.

Table 1 – Summary of levels of driving automation adapted from (SAE International 2016)

The six levels of driving automation are displayed relative to the roles of the human driver and automation (i.e. system) in the DDT and the DDT feedback, and to the ODD. The comparison to the BASt and NHTSA taxonomies is indicated in the last two columns.

SAE Level	Name	Definition	DDT		DDT fallback	ODD	BASt level	NHTSA level
			Sustained lateral and longitudinal vehicle motion control	OEDR				
Driver performs part or all of the DDT								
0	**No driving Automation**	The performance by the driver of the entire DTT, even when enhanced by active safety systems	Driver	Driver	Driver	N/A	**Driver only**	**0**
1	**Driver Assistance**	The sustained and ODD-specific execution by a driving automation system of either the lateral or longitudinal vehicle motion control subtask of the DDT (but not both simultaneously) with the expectation that the driver performs the remainder of the DDR	Driver and system	Driver	Driver	Limited	**Assisted**	**1**
2	**Partial Driving Automation**	The sustained and ODD-specific execution by a driving automation system of both the lateral or longitudinal vehicle motion control subtask of the DDT with the expectation that the driver performs the remainder of the DDR	System	Driver	Driver	Limited	**Partially Automated**	**2**
ADS ("System") performs the entire DTT (while engaged)								
3	**Conditional Driving Automation**	The sustained and ODD-specific performance by an ADS of the entire DDT with the expectation that the DDT fallback-ready user is receptive to ADS-issued requests to intervene, as well as to DDT performance-relevant system failures in other vehicle systems, and will respond appropriately	System	System	Fallback-ready user (becomes the driver during fallback)	Limited	**Highly Automated**	**3**
4	**High Driving Automation**	The sustained and ODD-specific performance by an ADS of the entire DDT and DDT fallback without any expectation that a user will respond to a request to intervene	System	System	System	Limited	**Fully Automated**	**3/4**
5	**Full Driving Automation**	The sustained and unconditional (i.e., not ODD-specific) performance by an ADS of the entire DDT and DDT fallback without any expectation that a user will respond to a request to intervene	System	System	System	Unlimited	-	

2.2.2 Motivation for vehicle automation

A key element to understand the context of vehicle automation is the motivation for the development of automated driving systems. Some of the main benefits expected from vehicle automation that motivate different stakeholders within the road transport community to pursue vehicle automation are:

Improvement of road safety

Vehicle automation is assumed to have the potential to improve road safety by reducing the number of accidents. Some studies have shown that 90-95% of the road crashes are attributed to the human driver error (Treat 1977; NHTSA 2008, 2015); it is assumed that automation will remove causal factors like driving under the influence of alcohol, speeding, distraction, human perception failures, etc. and therefore eliminate the crashes due to the human error.

Gain of comfort and time for the driver

A second expected benefit is the gain of comfort and time for the driver that results from automation taking over the driving task. Driving in cities with heavy traffic can be a very unpleasant experience that causes stress and fatigue to drivers. Low and intermediate levels of automation can reduce a driver's workload by taking over some of the driving tasks (M. S. Young and Stanton 2007). Moreover, high levels of automation can allow the driver to gain time by using the time that was previously dedicated to driving activities, to perform secondary tasks such as replying emails, reading, watching a film, etc.

More efficient fuel and energy consumption

Automated driving is expected to reduce the unnecessary acceleration and deceleration, to mitigate congestion, to improve fuel consumption and to lower the carbon dioxide emissions (Manzie, Watson, and Halgamuge 2007; Brown 2013).

Intelligent traffic management

Vehicle automation has the potential to improve traffic flow and to optimize the use of the traffic space. This can help to mitigate congestion and to Increase highway capacities by allowing vehicles to be closer to each other.

Mobility for everyone

Finally, vehicle automation is also expected to bring mobility to everyone, including to people that cannot drive a vehicle such as elderly people, people with visual impairments, people that are too young to drive, etc. or people that have no access to mobility (Alessandrini et al. 2015).

2.2.3 Paths to vehicle automation

Two paths to vehicle automation can be distinguished. While the first path referred to as "something everywhere" is being embraced by traditional vehicle manufacturers; the second path called "everything somewhere" is being followed by newcomers such as Google and Uber.

"Something everywhere" path

The first path consists of equipping conventional vehicles with increasingly sophisticated driving systems which perform parts or all of the entire driving task under several conditions. Highways tend to be the earlier road environment for conditional and high driving automation (SAE levels 3 and 4) due to the uniform design of the roads and their simpler interactions in comparison with the interactions of urban environments.

Some examples of the systems on the firs path are:

- Traffic Jam Assist: This system performs lateral and longitudinal vehicle motion on highways at slow speed in congested conditions.
- Highway pilot: This system performs lateral and longitudinal vehicle motion on highways with a speed range from low to high speeds. Some systems also perform lane change and respond to merging traffic.
- Automated valet parking: This system enables the driver to depart the vehicle at a parking garage entrance and to instruct the vehicle to park itself. Additionally the driver can also summon the vehicle to exit the parking area.
- Automated vehicle platoons: This system enables to perform lateral and longitudinal vehicle motion of several vehicles (e.g. trucks) which are closely spaced and tightly coordinated. A driver may be present in the vehicle leading the platoon.

"Everything somewhere" path

The second path consists of developing nonconventional vehicles which can perform the entire driving task without a human driver, firstly in limited contexts and the gradually expanding the range and the conditions; this path involves exclusively full driving automation (SAE level 5). The early use cases of this path include passenger shuttles and taxis that operate in central business districts, airports, universities, hospitals and other semi closed environments.

2.2.4 Challenges for vehicle automation

Some of the major challenges for the development and deployment of automated vehicle systems in terms of technology, legal and regulatory frameworks, and road user's interaction with automation, are described below.

Technical challenges for vehicle automation

Despite the fact that technologies such as Advanced Driving Assistance Systems (ADAS) and V2X communication represent a starting point for vehicle automation (also referred to as the building blocks), there are still numerous technical challenges to overcome before achieving full vehicle automation. For instance, the improvement of vehicle sensor capabilities and detection algorithms, the integration of multisensory systems and extending data fusion algorithms are necessary to achieve a robust perception on the driving environment.

Moreover, automated driving systems need a proper understanding of the spatio-temporal relationships of the vehicle and its environment and predicting the likely behavior of the entities sharing the same workspace of the vehicle (Eskandarian 2012). Regarding localization, overcoming the estimation errors of global coordinates like GPS, fusion of data from GPS receivers with other sensors data, and improving the quality of digital maps and map-matching algorithms, remain key issues to solve the problem of determining the position of the vehicle with respect to the environment (Pendleton et al. 2017).

Additionally, the advances in vehicle control tested in simulation need to be tested under real conditions to ensure that the automated driving system follows the intention of the higher-level decision-making processes (Pendleton et al. 2017). Also, the increasingly complexity of software brought by more code, more conditional branches, high-dimensional interfaces, and complex (often novel) algorithms, requires balancing computational resource

allotments and finding new verification methods (Koopman and Wagner 2016). To the previous challenges it must be added that the available solutions need to be at an acceptable level of costs. Even If the technology for vehicle automation is developed, it has to be available at a reasonable cost for it to be implemented in production vehicles.

Regulatory and legal challenges for vehicle automation

The main regulatory and legal challenge for automated vehicles is the need to assess the existing traffic and vehicle regulatory and legal frameworks (which were defined for conventional vehicles) relative to vehicle automation. For example, at the international level, the Vienna Convention on Road Traffic of 1968 established a set of traffic rules among the contracting parties, including an Article 8 in which a vehicle shall have a driver who is always fully in control and responsible for the behavior of a vehicle in traffic (UN 1968). Since automation provides lateral and longitudinal control of the vehicle while the automated driving system is engaged, it was not clear whether or not automated vehicles were legal relative to the Vienna Convention. Recently, the Vienna Convention was updated to allow transferring driving tasks to the automated systems as long as the driver can override or switch off the system (UNECE 2016). Nevertheless, the amended convention still demands that every vehicle must have a driver.

Additionally, a second major regulatory update under discussion is the introduction of technical provisions for self-steering systems. Moreover, at the European level, the technical requirements for motor vehicles and type-approvals in the European Union to ensure that new vehicles on the market provide a high level of safety and environmental protection, may need to be adapted for the higher levels of automation (Pillath 2016). At the national level, national governments must adapt traffic rules and driver education and licensing regulations. Further, the regulatory framework for automated driving trials on open roads also needs to be defined.

Liability challenges for vehicle automation

For traditional vehicles it is clear that the human driver is liable for the harm to persons and property resulting from a crash (unless there is a technical failure of vehicle malfunction involving product liability or failures in road design and maintenance) (B. W. Smith 2016).

The delegation of driving tasks to automation also shifts the responsibility of driving to automation and therefore automation (and manufacturer) could be liable for a crash. Consequently, the liability regimes might need to evolve in order to fairly determine the responsibilities of the driver, automation, automakers, suppliers, etc. Additionally, the insurance industry will also have to adapt to these liability challenges and to the decrease of liability policies as vehicle automation is expected to cause fewer crashes.

Human Factor challenges

The study of human factor challenges related to vehicle automation has mainly focused on the intermediary levels of automation (NHTSA 2014; Natasha Merat and de Waard 2014; Natasha Merat et al. 2014). This does not mean that full automation does not face human factor issues, however, the shared-control, need for fallback users in case of malfunctions or takeover requests, and the fact that automakers are first developing SAE level 3 and 4 systems, has resulted in more efforts being placed on intermediary levels.

There are human factor challenges related to the driver's understanding of the system such as mode confusion and authority issues. For example, when the human driver transitions between several levels of automation, s/he can experience mode confusion. Further, there is a risk that the human driver perceives automation as the ultimate controlling authority when the human driver is the controlling authority and vice versa. This might lead human drivers to misunderstand their responsibilities. Moreover there can be error handling, in which the automated system may not correctly account for the intentions of the driver and may act inconsistently with the expectations of the driver (NHTSA 2014).

Additionally, automation can also affect driver's situation awareness which is the perception of the elements in the environment within an environment of time and space, the comprehension of their meaning and the projection of their status in the near future (Endsley 1995; Kaber and Endsley 1997; N. Merat and Jamson 2009; Natasha Merat et al. 2014). A deterioration of the situation awareness can be very problematic if drivers have to retake control within a short period of time (de Winter et al. 2014). Additionally, automated systems may also lead drivers to boredom which can cause distraction.

Inappropriate trust includes misuse in which user violates critical assumptions and relies on automation inappropriately, disuse in which the user rejects automation, and abuse of

automation in which designers introduce an appropriate application of automation (Bainbridge 1983; Parasuraman and Riley 1997). Lastly, if automation is as reliable and useful, drivers may rely heavily on automation and fail to utilize their own skills, leading to driver's skill degradation.

2.3 Road safety

Section 2.3 provides two definitions for road safety: road safety as the lack of safety and road safety as a system. Next, it presents the road safety perspectives over time; from seeing crashes as random events in 1900 to today's view in which crashes are seen as the result of the system not being well adapted to the road user. Finally, the Safe System approach to road safety is described.

2.3.1 Road safety as a lack of safety

Road safety is often defined using the quantitative measures related to road trauma such as the number of fatalities or injured persons in a unit of time. These measures nearly always focus on the magnitudes of departures from a total absence of some type of harm, rather than directly on safety as such (Evans 2004). Consequently, we look at the lack of safety instead of safety itself and end up referring to road safety as the number of fatalities or injuries resulting from traffic crashes (Elvik 2009; Risto Kulmala 2010).

Nilsson introduced a conceptual framework that helps to describe and model the road safety situation in three dimensions—Exposure, Risk and Consequence— and estimate the number of fatalities and injuries (Nilsson 2004). Exposure usually refers to the amount of travel in which crashes may occur, expressed as the number of vehicle kilometers performed or hours travelled. Risk denotes the probability of a crash which is often expressed as crash rate. Consequence is the probability of fatality (or a particular level of injury).

The three dimensions have a multiplicative relationship with regard to safety (Nilsson 2004):

$$\text{Traffic safety problem} = \text{Exposure} \times \text{Risk} \times \text{Consequence}$$

$$\text{Number of injured} = \text{Exposure} \times \left(\frac{Crashes}{Exposure}\right) \times \left(\frac{Injured}{Crashes}\right)$$

$$\text{Numer of fatalties} = \text{Exposure} \times \left(\frac{\text{Injured}}{\text{Exposure}}\right) \times \left(\frac{\text{Fatalities}}{\text{Injured}}\right)$$

This relationship is further illustrated in figure 4, where the volume of the rectangle is the expected number of injured or fatalities.

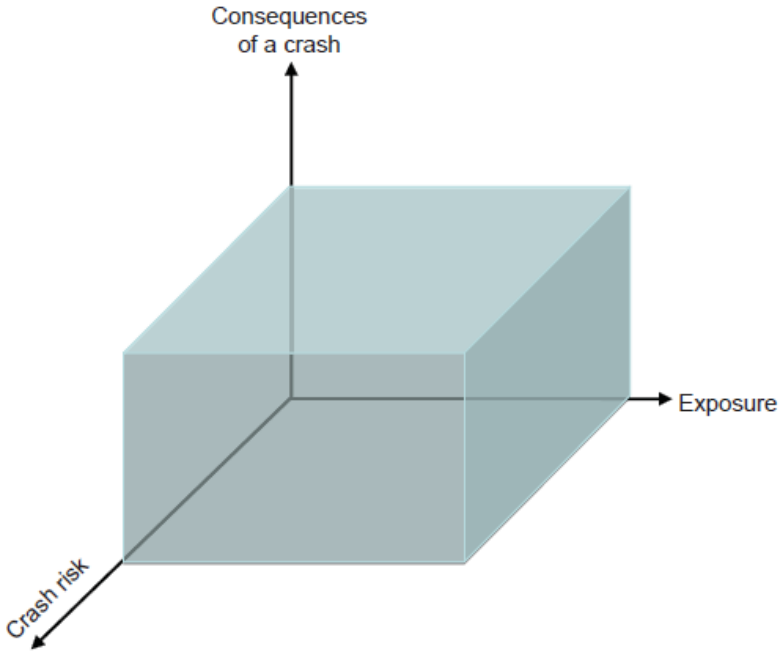


Figure 4 – The three dimensions of road safety

2.3.2 Road safety as a system

While it is useful to define road safety terms of traffic “unsafety”, this thesis considers an alternative and broader definition by integrating three basic components: road safety prevention, road safety analysis and the current state of road safety, into a system composed of multiple interacting stakeholders that aims to improve the level of safety on the roads via safety measures. In order to achieve this aim, stakeholders participate in a continuous process (displayed in figure 5) in which road safety prevention defines and implements measures to target safety problems that have been identified by road safety analysis, which in turn uses information on the state of road safety e.g. the number of crashes, fatalities and injuries, and safety performance indicators.

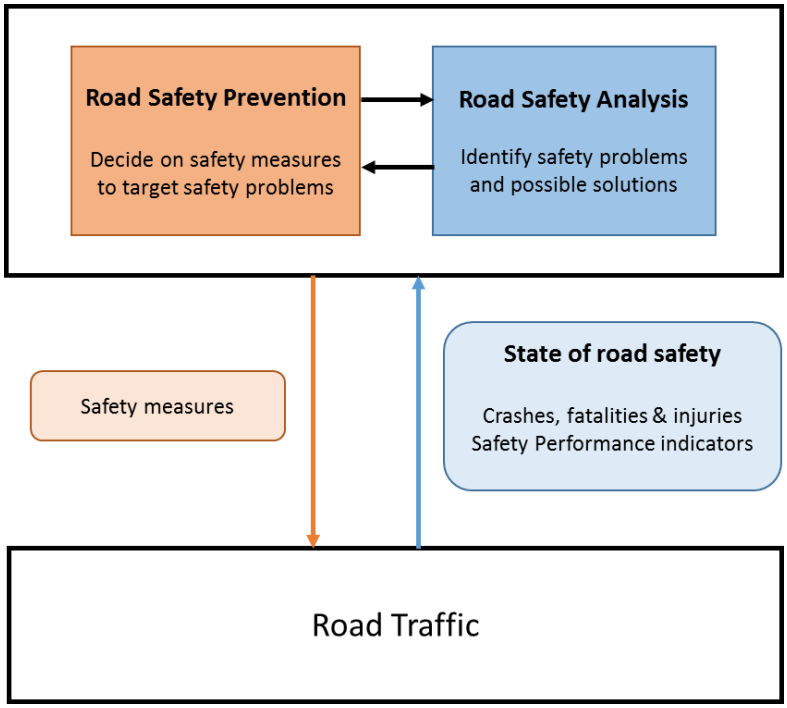


Figure 5 – The process to improve road safety

Road safety prevention

The ultimate goal of road safety is to improve the level of safety of the road transport system i.e. to reduce the number of persons being killed, seriously injured or involved in a crash. This goal is achieved through road safety prevention, which involves the elaboration of prevention strategies, and the selection and implementation of effective safety interventions. Therefore, prevention becomes the fundamental component of road safety.

As seen in figure 5, road safety prevention uses the information on safety problems and possible solutions identified by road safety analysis (along with other information such as feasibility, costs and acceptability) to develop strategies and to define safety measures that aim at solving the problems. Although the elaboration of these strategies is often seen as a matter of the government, it can also be a matter of private stakeholders. For instance, automakers can contribute to prevention by complying with the safety regulatory framework, by participating in consumer tests that evaluate safety such as the Euro NCAP rating system, and finally by a brand differentiation centered on safety.

Road safety analysis

Road safety analysis uses the data on the current state of road safety and the data obtained from studies (e.g. epidemiological studies, naturalistic driving studies, test track studies, simulator studies, etc.) to identify safety problems and possible solutions generally divided into risk factors such as alcohol and distraction, risk groups like pedestrians and two wheelers, and types of crashes.

Stakeholders such as research institutes, government agencies and automakers use various disciplines like statistics, engineering, medicine, psychology, sociology and ergonomics, to analyze road traffic and to identify problems and possible solutions. For instance, epidemiological studies based on statistics determine the incidence, prevalence, risks and relative risks of different risk factors e.g. alcohol, excessive speeds, inattention, etc. (Elvik 2013), of different user groups like passenger vehicles, two wheelers, etc. (Vlahogianni, Yannis, and Golias 2012) and of different risk groups such as young drivers, and pedestrians. Moreover, in-depth crash analysis looks into detail of crashes, case by case, in order identify accident mechanisms and new research areas (Van Elslande 2000a). Questionnaires and interviews are also conducted to evaluate drivers' attitudes toward risk and vehicle systems (Assailly 1993).

State of road safety

The state of road safety cannot be accessed directly; the data on reported crashes, fatalities, and injures are employed to represent the current safety level on the roads. More recently, it has been argued that the number of crashes is not enough to understand the process that leads to accidents and thus safety performance indicators (SPI) have been suggested as a complementary alternative. The SPI are measures related to behavioral characteristics (e.g. speed and rates of drinking and driving), to infrastructure (e.g. pavement friction), vehicles and trauma (Hakkert, Gitelman, and Vis 2007). These data are used by road safety analysis to identify problems. Additionally, these data also serve to monitor the changes in road traffic and the effects of past safety measures.

2.3.3 Road safety perspectives over time

Over time, there have been various perspectives that have tried to explain road crashes; they have influenced road safety analysis and prevention in research and practice. The road safety perspectives have been classified through time periods by several authors (OECD 1997; Elvik 2009; Hagenzieker, Commandeur, and Bijleveld 2014); although there are some minor discrepancies over the time interval for each period, the following views can be found in all the classifications.

Crashes as random events (1900-1920)

In the early days of motorization accidents were seen as a random event over which humans had no control and therefore being involved in an accident was just a matter of bad luck. The Poisson distribution was used to model and thus describe the random process that led to accidents. At this point, research was focused on collecting basic statistics and answering the question of "What" happened in accidents.

Crashes are caused by crash-prove drivers (1920-1950)

Abnormal concentrations of accidents showed that some people had more accidents than others and that it could not be explained by randomness alone, leading to the assumption that some people were more prone to have accidents than others. Research was focused on identifying "Who" were particularly prone to crashes. This shifted the paradigm of road safety from believing that crashes were a random event to believing that a few drivers who were more prone to accidents were responsible for crashes.

Crashes are mono-causal (1940-1960)

The growth of motorization brought an increase in the number of crashes, which proved that crashes could happen to everyone and not just a few people. It was proposed that the only way prevention was possible, was by finding the real causes of accidents; as a consequence, research was focused on "How" crashes happened and finding the single or root cause of crashes. In-depth case studies were conducted to find the actual causes of accidents. One of the main findings was that 85-90% of the crashes were caused by human driver errors.

Crashes are multi-causal (1950-1980)

The developments of systems theory and epidemiological theory helped researchers to realize that crashes were not due to a single cause but to the combination of factors that contribute to the occurrence of a crash; crashes are the result of maladjustments in the multiple interactions between complex systems; it is not possible to pick one of the parts of the road transport system as more crucial than the others for its successful operation. The solution consisted in modifying the technical components of the system (vehicles and roads) such as vehicle design.

Crashes are the result of behavioral influence (1980-2000)

This period was highly influenced by behavioral theories that argue that the human risk assessment and human risk acceptance are very important determinants of the actual number of accidents in an activity. According to this perspective, crashes are closely related to the risk that a road user is willing to endure. The road user is seen as the weak link and consequently the only way of lowering the number of crashes is by changing the target level of risk (desired level of safety) of a society.

Crashes are the result of the system not being well adapted to the road user (1990-now)

Nowadays, crashes are seen as inevitable phenomena due to the fact that the road transport system is not well adapted to the road users. This view denotes a new paradigm shift that has been called "Safe Systems approach", in which the solution to improve road safety is to adapt the system to the psychological and physical conditions of the human beings and to share the responsibility of road safety between all the stakeholders of the road transport system and not the road user alone. It focuses on the better implementation of existing policies and a systems management perspective.

2.3.4 Safe System approach

The International Transport Forum (ITF) recently encouraged the governments of all countries to use the United Nation Sustainable Development Goal that sets a target to halve the number of road fatalities and serious injuries by 2020, to review their road safety policies and to explore the Safe System approach to road safety (ITF 2016).

The Safe System approach embodies the long-term policies that have been adopted by countries with leading road safety improvement results such as Vision Zero in Sweden (Tingvall 1995) and Sustainable Safety in the Netherlands (Wegman, Aarts, and Bax 2008).

The Safe System approach is based on the ethical imperative that no human should be killed or seriously injured on the roads; it aims to develop a road transport system that is better able to accommodate human error and take into account the vulnerability of the human body. The Safe System approach is based on four principles.

Principles:

1. Road users make mistakes that can lead to crashes:

Humans cannot be faultless road users throughout all the time; even if their intention is to drive in a safe manner at all times, road users make mistakes that can lead to road crashes. Road user's human error should no longer be considered as the primary cause of crashes but as a consequence of latent failures caused by the actors at the sharp end of the road transport system. Therefore, the capabilities and limitations of road users must be considered in the design and operation of the road transport system.

2. Limited physical crash tolerance:

The human body has a limited physical ability to absorb the kinetic energy a crash exerts before harm occurs. Consequently, the reduction of operating speeds can mitigate the risk of injury and the risk of common road user's errors.

3. A shared responsibility for road safety:

While road users have a responsibility to comply with traffic regulations and to drive in a safe manner, the actors involved in the design and operation of the system, need to accept and share the responsibility for the safety of the system.

4. Strengthen all parts of the system:

All the layers of the system (i.e. design and operation of road infrastructure, operating speeds, vehicles and human behavior) must be strengthened and managed holistically in order to multiply their effect, so that the combination of the layers cover for each other in case one element fails.

A comparison between the traditional road safety policy and the Safe System approach is displayed in table 2 (ITF 2016). It illustrates the main differences of the two approaches in terms of safety problem, goals, planning and approaches, causes of the problem, responsibility and system interactions.

Table 2 – Comparison between the traditional road safety policies and the Safe System approach (ITF 2016)

Comparison criterion	Traditional road safety policy	Safe System
What is the problem?	Try to prevent all crashes	Prevent crashes from resulting in fatal and serious causalities
What is the appropriate goal?	Reduce the number of fatalities and serious injuries	Zero fatalities and serious injuries
What are the major planning approaches?	Reactive to incidents Incremental approach to reduce the problem	Proactively target and treat risk Systematic approach to build a safe road system
What causes the problem	Non-compliant road users	People make mistakes and are physically fragile/vulnerable in crashes. Varying quality and design of infrastructure and operating speeds provides inconsistent guidance to users about what is safe use behavior
Who is ultimately responsible?	Individual road users	Share responsibility with system designers
How does the system work?	Is composed of isolated interventions	Different elements of a safe system combine to produce a summary effect greater than the sum of the individual treatments – so that if one part of the system fails other parts provide protection

The conceptualization of the Safe System:

The interactions among the several layers, actors, activities and components of a Safe System are illustrated in figure 6. The four principles of a Safe System interact with the design and operation of the road transport system. The center of the figure displays the first principle, that is, physically vulnerable road users can make mistakes leading to crashes. The second layer shows the relationship between speed, roads, roadsides and vehicles, which should support road users to behave safely in traffic and ensure that when a crash does occur, it does not result in serious injuries or death. The third layer, captures the second principle i.e. the human body has a limited physical crash tolerance.

In a Safe System, the components of the second layer and their interactions must be managed to avoid crashes that exceed the level of kinetic energy a human can absorb before

harm. The fourth layer represents the post-crash response and medical care which provides emergency care for crashes with impact forces that cause serious physical harm. The second and third layers capture the fourth principle, in which all the parts of the system must be strengthened to multiply their effects, in case one part fails. Lastly, the fifth layer illustrates the third principle of shared responsibility, in which all the actors of the system must work together to prevent crashes resulting in serious injury or death. Furthermore, this principle can be supported by a management by objectives, to provide data and results that reinforce collaboration between all actors.

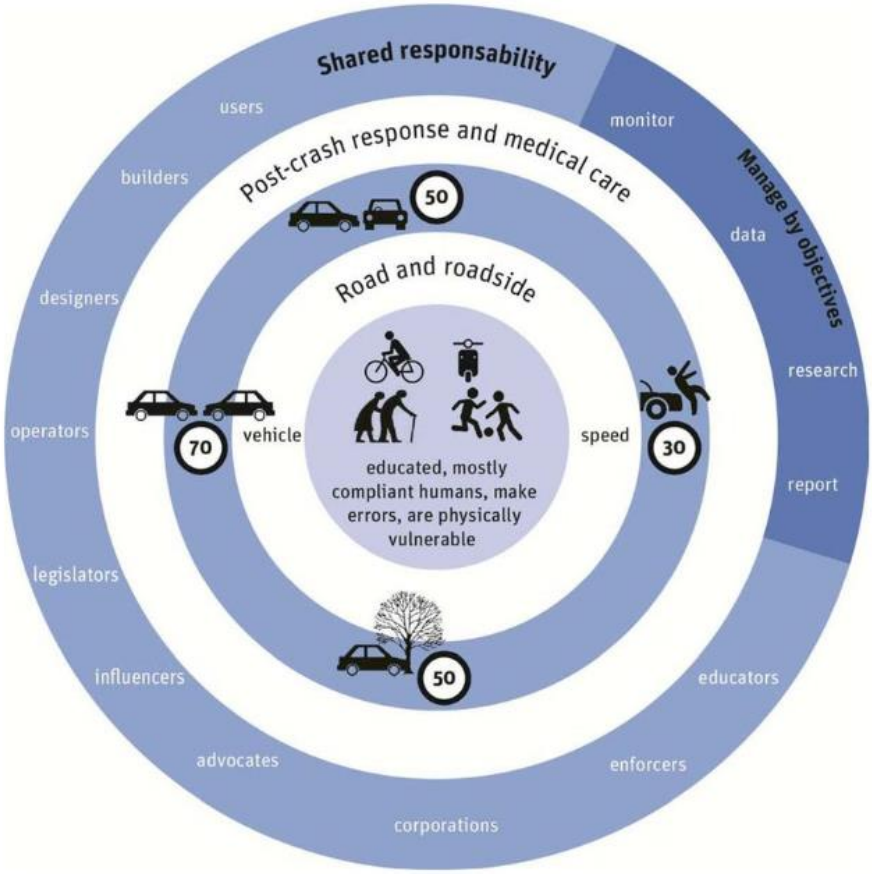


Figure 6 – Conceptualization of the Safe System (ITF 2016)

2.4 Systems theoretic approaches to safety

Section 2.4 describes the studies from the literature that suggest moving towards a systems theory approach to road safety. Next, it presents the three most popular systems theoretic approaches namely Rasmussen's Risk Management Framework, Leveson's Systems-Theoretic Accident Model and Processes (STAMP) and Hollnagel's Functional Resonance Analysis Method (FRAM). Lastly, it provides a synthesis of the three approaches.

2.4.1 Systems theory and road safety

The number of road traffic deaths globally has plateaued at 1.2 million a year since 2007 (World Health Organization 2015). Moreover, road crashes are the current leading cause of death among people aged 15 to 29 years. Further, road traffic injuries are the ninth leading cause of death across all age groups and are predicted to become the seventh leading cause by 2030 (World Health Organization 2015). While the risk of road traffic death is the highest in low and middle-income countries (notably in the African region where the road traffic death rate is 26.6 per 100000 population in 2013); even the high-income countries that have adopted a Safe System approach such as Sweden and the Netherlands, which have the lowest road fatality rates (2.8 and 2.4 per 10000 population in 2013), are having a slow decrease or stagnation of the number of road traffic deaths since 2010 (European Commission and DG for Mobility and Transport 2015).

As a result of this unacceptable high level of road trauma and its small improvements, road safety researchers and practitioners have questioned the need for a paradigm shift towards systems theory (Larsson, Dekker, and Tingvall 2010; Salmon, McClure, and Stanton 2012; Salmon and Lenné 2015). In fact, the road transport system is an increasingly complex sociotechnical system comprising many inter-related components and complex interactions that go beyond the driver-vehicle-environment compound, which could use approaches from systems theory to cope with complexity and to support crash accident analysis and prevention (Salmon, McClure, and Stanton 2012).

(Hughes et al. 2015) reviewed and evaluated 121 models relevant to road safety from a wide variety of fields e.g. transport, occupational safety, food industry, education and health, and recommended that the models from systems theory should be comprehensively applied in road safety research and practice at all levels, notably at the whole system level. In the model evaluation, they categorized the models into 6 types (component models, sequence models, intervention models, mathematical models, safety management models and systems models) and compared their potential to be adapted to the road transport system relative to four criteria: model use, strengths, weaknesses and relevance to road safety. According to the evaluation, the systems models have been used to analyze systems, the effects of countermeasures, influences and consequences. Moreover, the identified strengths were: their assistance in the understanding of the whole system and the

contribution of components, their consideration of the holistic outcomes and interdependencies among components, their inclusion of the structural, human resource and symbolic aspects, and their theoretical basis. The weaknesses included the difficulty to apply systems models on systems with many components and multiple relationships which makes the process too complex to analyze, and the difficulty to analyze and provide strong quantitative evidence. Lastly, the descriptive models and detailed analytical models provided in systems models were found to be particularly relevant to road safety.

Furthermore, (Larsson, Dekker, and Tingvall 2010) and (Hughes, Anund, and Falkmer 2015) demonstrated that although the road safety strategies from the Safe System mention some of the main features of systems theory, these features have not been comprehensively included in such strategies. (Larsson, Dekker, and Tingvall 2010) evaluated the road-user focused approach (i.e. the traditional road safety approach) and the vision zero approach (i.e. the Safe System approach) relative to three key features of systems theory: safety as an emergent property, system component performance variability and systems as hierarchical structures. Their findings showed that the three key features from systems theory were not present in the road-user approach and that even though the vision zero approach takes a step towards systems theory, there is still room for articulating more features of systems theory. Additionally (Hughes, Anund, and Falkmer 2015) compared five recent road safety prevention strategies (including vision zero) which fall under the Safe System approach, and examined them with respect to their foundations in systems theory and safety models. Their results confirmed that these modern strategies do not comprehensively include essential aspects of systems theory such the interdependencies between their basic components. To conclude, they recommended that further development is needed to completely apply the concepts of systems theory in road safety strategies.

Moreover, the first applications of systems theory to road safety have illustrated the potential of these approaches to contribute to a better understanding of the road transport system and the factors that interact to create road trauma. For example (Scott-Parker, Goode, and Salmon 2015) demonstrated the utility of systems theory approaches to depict the current knowledge on the multiple actors, contributing factors and countermeasures related to young driver safety at all levels of the road transport system. Furthermore (K. L. Young and Salmon 2015) analyzed the existing knowledge on driver distraction regarding the

responsible actors, the enablers of driver distraction across the different levels of the road transport system and speculated what a systems theory approach on driver distraction might entail. Finally (Salmon, Read, and Stevens 2016) showed that a systems theory approach can be used to model the actors of the road transport system and the control and feedback mechanisms between them as a hierarchical control structure.

To conclude, the need to move towards a systems approach to road safety stated in the literature and the promising results of the first applications, suggest that the models and methods based on systems theory could constitute a suitable conceptual framework to address the three research questions of this thesis. Among the multiple models and methods that have been described as being based on systems theory (Underwood and Waterson 2013) confirmed that the three most popular models referred to as systemic models i.e. the Risk management Framework, the System-Theoretic Accident Model and Process (STAMP) and the Functional Resonance Analysis Method (FRAM), do include key systems theory characteristics: system structure, system component relationships, and system behavior. Consequently, these three models were reviewed to determine which model could provide the conceptual framework for the thesis.

2.4.2 The Risk Management Framework

The risk management framework developed by Rasmussen is the culmination of a research program that he started in the late 1960's and ended in the late 1990's. It represents a shift from a microscopic-individual to a macroscopic-sociotechnical view of safety (Le Coze 2013; Waterson, Le Coze, and Andersen 2017). In the macroscopic view, Rasmussen argued that the present dynamic society has brought changes into systems (e.g. a very fast pace of change of technology, the increase of potential for large-scale accidents, an aggressive and competitive environment, etc.) which make it necessary to transform the way risk management is modeled (J. Rasmussen 1997; Rasmussen & Svedung 2000). Accordingly, the Risk Management Framework sees risk management as a control problem embedded in an adaptive socio-technical system, and models the system behavior based on functional abstraction instead of structural decomposition. Rasmussen developed two complementary models to support his risk management framework: the socio-technical system model and the model of migration.

The socio-technical system model

As seen in figure 7, this model represents the socio-technical system (STS) as a flow of information between six levels that interact in a hierarchical way, while being opened to environmental stressors like the political change and public awareness, the market conditions and financial pressure, the competency levels of education and the fast pace of technological change. The levels of politicians, managers, safety, officers, and work planners manage safety through laws, rules, and instructions that aim to control the hazardous physical process. Additionally, these levels aim to motivate workers and operators, train them, guide them, or constraint their behaviors through rules, and equipment design. Even tough, traditionally, each level is studied individually (or horizontally) by a particular research discipline (e.g. at the top level, the government regulates safety through the legal system and is studied by political science), Rasmussen points out the critical importance of studying the vertical interactions among these levels in a cross-disciplinary fashion.

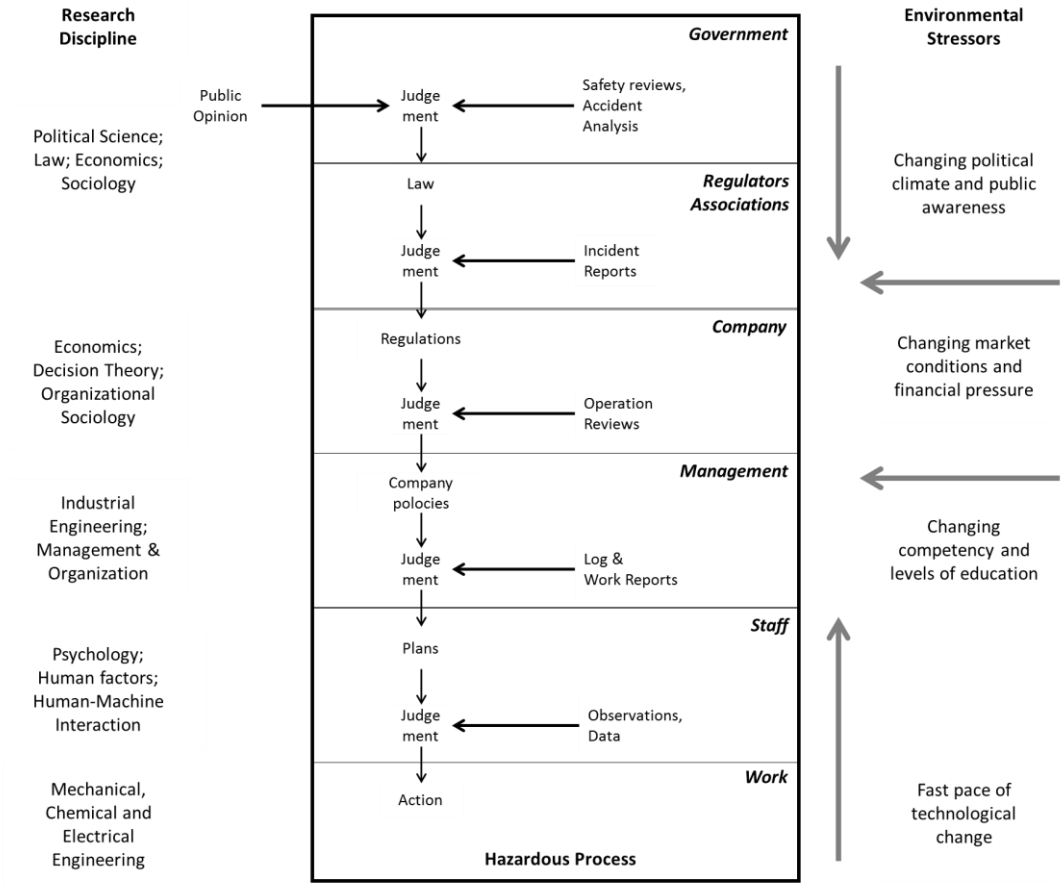


Figure 7 - Model of the socio-technical system (Rasmussen 1997)

Model of migration toward the boundary of acceptable performance

The model of migration or the dynamic model of safety and system performance (figure 8), illustrates how economic considerations and workload pressures can move the system away from safe performance and closer to the margin of error. In this model, Rasmussen proposes to represent the system behavior by focusing on the behavior of operators in the actual, dynamic work context. The model considers that there is a safety space of performance delimited by boundaries such as individual unacceptable workload, financial and economic constraints and perceived acceptable performance, within which operators can navigate freely. However, gradients towards least effort and pressures towards efficiency induce variations in human behavior analogous to "Brownian movements", which make the system migrate towards the boundaries of acceptable performance and safety performance, and may ultimately lead to an accident if control is lost at the boundaries.

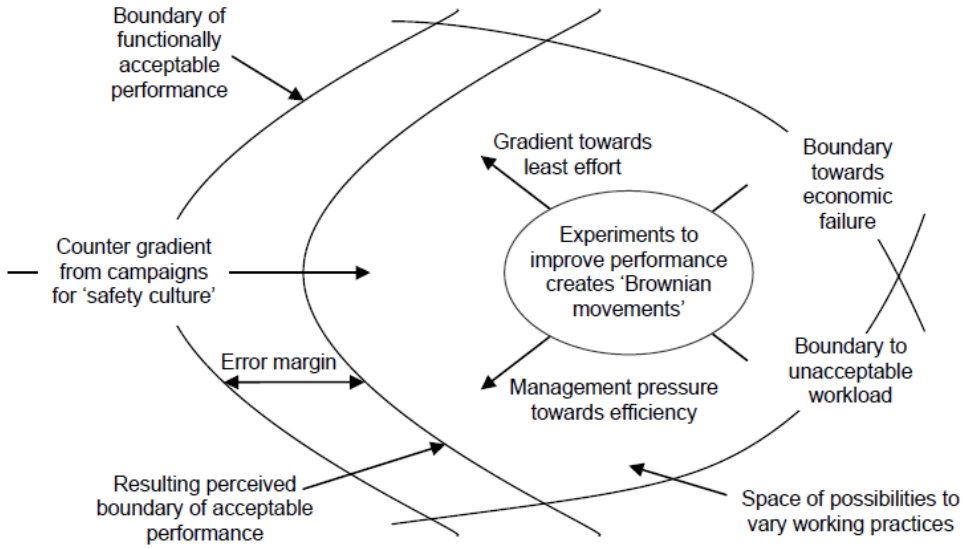


Figure 8 - Model of migration toward the boundary of acceptable performance (Rasmussen 1997)

AcciMap

The Risk Management framework was extended into an accident analysis analytical tool called AcciMap, which intended to support proactive risk management (Rasmussen & Svedung 2000); however, in practice AcciMaps are mostly used for retrospective accident analysis (Underwood 2013).

Based on the idea that safety is influenced by all the sociotechnical levels, AcciMap is a multilayered diagram that maps and models the events, actors, acts, and decisions involved in an accident and their interactions across the six sociotechnical levels, to identify the multiple factors that contributed an the accident. As observed in figure 9, an AcciMap contains nodes and arrows across the six sociotechnical levels that represent the causal flow of events and system states leading to an accident.

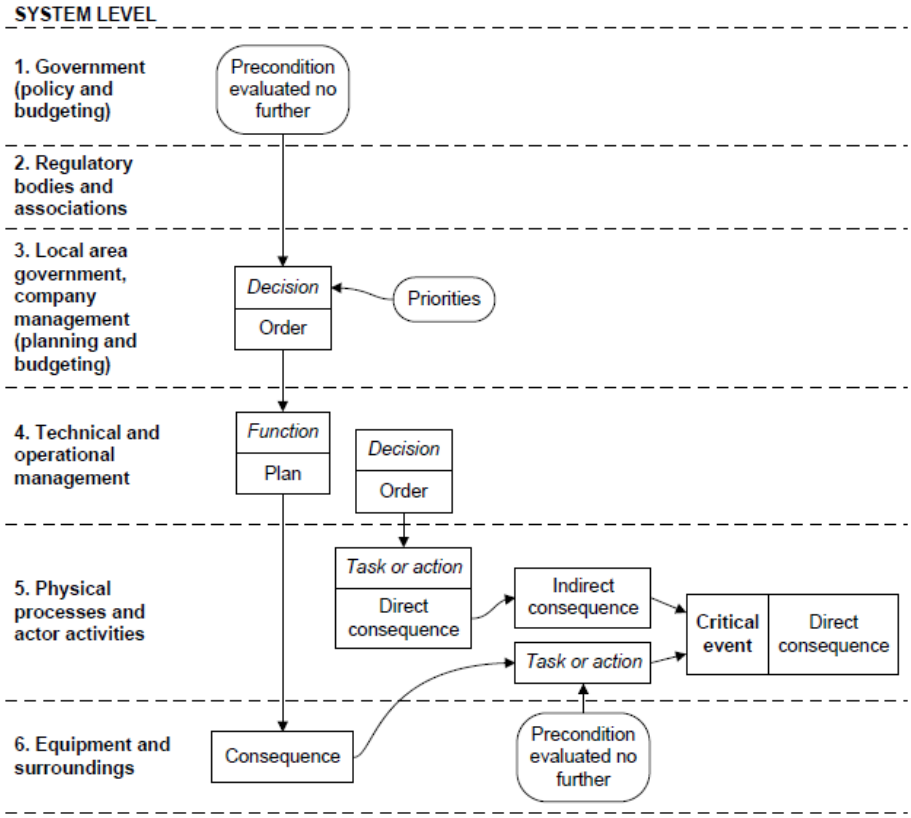


Figure 9 – AcciMap diagram (Rasmussen & Svedung 2000)

2.4.3 System-Theoretic Accident Model and Processes (STAMP)

Like Rasmussen, Leveson argues that the context and the changes in the systems being built have created a need for new accident models. As a result, Leveson proposed a new accident model called STAMP which is based on systems theory rather than reliability theory (Leveson 2004, 2011). STAMP sees safety as a control problem managed by a control structure embedded in a sociotechnical system which enforces a set of safety constraints on the system behavior (Leveson 2004, 2011). Accordingly, accidents are viewed as a loss of

control[5] that arises when external disturbances, component failures, or dysfunctional interactions among system components violate safety constraints i.e. when constraints are not adequately imposed or enforced on the system.

As observed in figure 10, in STAMP accidents occur when the system gets into a hazardous sate, which in turn occurs due to the inadequate enforcement of safety constraints on the system behavior.

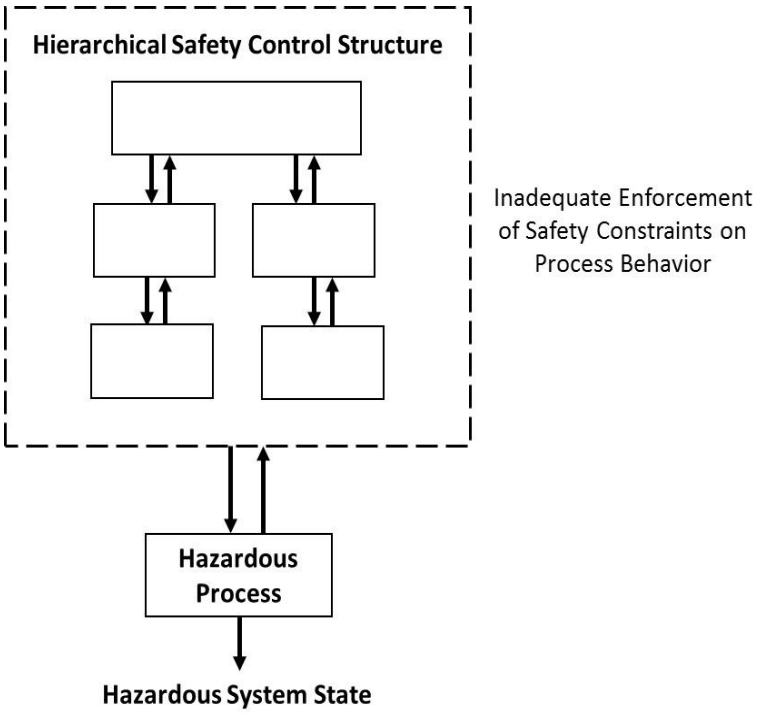


Figure 10 – Illustration of STAMP model (Leveson and Thomas 2013)

By seeing accidents as a dynamic control problem rather than a component problem, STAMP includes causal factors beyond component failure such as design errors, software requirement flaws, human behavior, unsafe interactions, migration of the overall system towards states of higher risks, etc.

Basic concepts

STAMP has three basic concepts: safety constraints, hierarchical control structures and process models. Safety constraints (which can be physical, human or social) are imposed on

[5] Control is a very broad term in STAMP; it entails physical, human and organizational controls.

the system to control system behavior and enforce safety. Inspired by Rasmussen's sociotechnical model displayed, Leveson represents the socio-technical system as multiple hierarchical levels with controllers and control processes operating at the interfaces between levels, where each level imposes constraints on the activity of the level beneath it (Leveson 2017b). Finally, the process models (displayed in figure 11) are the representations that every controller has about the system it is controlling, that help controllers make decisions regarding what control actions to provide to enforce safety constraints. Process models contain representations about the current state of the system and what the controller should do to control it. Additionally, the process models are kept up to date through feedback loops.

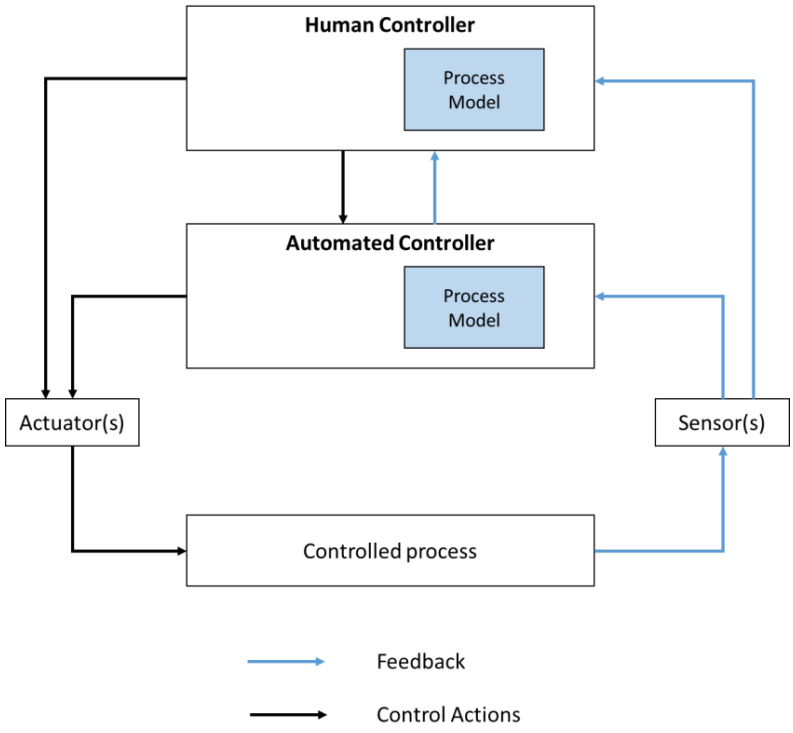


Figure 11 - Controllers and Process model

Systems-Theoretic Process Analysis (STPA)

STPA is a hazard analysis method based on STAMP which aims at identifying hazards and the scenarios leading to hazards so they can be eliminated or controlled before damage occurs. STPA analyses consist of identifying unsafe control action provided by the controllers of the system to define safety requirements (also called safety constraints); next, the identified

unsafe control actions are examined to elaborate scenarios leading to hazards and to define additional safety requirements.

Causal Analysis based on Systems Theory (CAST)

CAST is an accident analysis method with STAMP as its conceptual foundation. Consequently, it assumes accidents are caused by inadequate enforcement of safety constraints on system behavior. The main objective of CAST is to identify systemic factors that lead to accidents and to generate recommendations that eliminate or reduce unsafe behavior. CAST analyses involve examining the entire control structure (starting at the bottom and moving upward in the control structure) at the time of the accident to identify the violated safety constraints and unsafe behaviors at each level and the reasons why controllers behaved the way they did.

2.4.4 Functional Resonance Analysis Method (FRAM)

Hollnagel states that current sociotechnical systems are more or less tractable, that is, they have elaborate descriptions with many details, high rate of change, their principles of functioning are partly unknown and their processes are heterogeneous and possibly irregular. Consequently, Hollnagel argues that the methods based on simple linear thinking and complex linear thinking are not enough to comprehensible understand and model the dynamic and non-linear behavior of such systems. As an alternative, he proposes to use Resilience Engineering as a basis to a new method called FRAM (Hollnagel 2012).

According to FRAM, safety is compromised when the variability of the adjustments of everyday performance—which normally help things go right—, aggregates in unexpected ways and experiences functional resonance. The aim of FRAM is to look for and to monitor what is needed for everyday performance to go right in order to dampen the variability that causes unwanted outcomes and to amplify the variability that leads to wanted outcomes.

FRAM is built upon four principles:

1. **The equivalence of failures and successes:**

The assumption that failures happen because things go wrong leads to see failures and successes as having a different nature and to put almost all efforts in understanding why things go wrong. In FRAM, failures and successes are viewed as equivalent in the sense that they have the same origin; things go right and wrong for the same reasons.

2. The approximate adjustments:

Human performance is always variable due to a number of internal and external factors. The performance adjustments that humans make to match conditions are the reasons why everyday performance is successful and also why sometimes things go wrong.

3. Emergence:

This term is used to describe the occurrences that cannot be explained using the principles of decomposition and causality. Occurrences in which the effects are non-linear, where the final outcome might be due to transient phenomena or conditions that were only present at a particular time of time and space, as combinations that existed for a brief moment.

4. Functional resonance:

Stochastic resonance is used as an analogy to describe the relationships and dependencies among the system's functions and how everyday performance variability can lead to unexpected outcomes. However, Hollnagel refers to functional resonance rather than stochastic resonance given that variability is mainly due to the approximate adjustments of people, individually and collectively and of organizations.

FRAM as a method to analyze past and future events:

Based on these principles, Hollnagel proposed a four-step bidirectional method for both accident analysis and risk analysis. In the first step, the functions of the system are identified and described according to the six aspects: input, output, preconditions, resources, time and control displayed in figure 12.

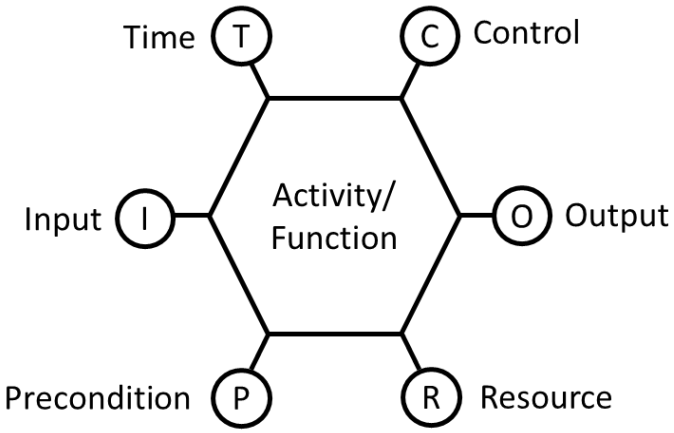


Figure 12 – The six aspects that describe a function in FRAM

In the second step, the potential for performance variability is characterized by categorizing the functions as technological, human or organizational, describing their sources of variability and the output variability in terms of time and precision or failure modes. Thirdly, the way how variability may be combined (aggregation of variability) is identified by analyzing the dependencies between the functions and the potential of unwanted resonant connections. Lastly, countermeasures necessary to manage and dampen function variability are proposed.

2.4.5 Synthesis of the systems theoretic approaches to safety

Table 3 illustrates a synthesis of the three aforementioned systemic models relative to their conceptual basis, view of accidents, main principles, modeling approaches and interesting features for the thesis. Regarding the conceptual basis, it is confirmed that the three models include concepts from systems theory. As expected for a model that was intendedly developed based on systems theory, STAMP has the strongest systems theoretic conceptual foundations. While there are some differences on the view of accidents, the three models consider accidents as the result of non-linear interactions which are not necessarily due to failures.

The principles of the Risk Management Framework and STAMP are rather similar, as Leveson points out *"My own attempts to extend Rasmussen's ideas to engineering practice have involved improving engineering specifications, particularly the requirements engineering process; creating a new, more powerful, model of accident causation that better explains the cause of accidents in human-operated, software intensive, sociotechnical systems; and creating new hazard analysis techniques that integrate humans into the generation of causal scenarios"* (Leveson 2017b). On the other hand, FRAM lists other principles such as equivalence of failures and successes and functional resonance; FRAM does not openly mention hierarchy.

The modeling approaches of the Risk Management Framework and STAMP are also comparable; they both represent the sociotechnical system as hierarchical levels. FRAM takes a different modelling approach by representing system's functions as hexagons and the dependencies between the functions. Finally, several interesting features for the thesis were identified in the three models. The three models seem to be able to model the road

transport system in a structured fashion which is assumed to facilitate the analysis of the system. Moreover, the methods based on the three models can guide hazard analysis and accident analysis. However, STAMP stands out among the three models due to its explicit interest on automation.

Table 3 – Synthesis of the three models

Characteristic	Rasmussen's Risk Management Framework	Leveson's STAMP	Hollnagel's FRAM
Conceptual basis	-It includes various concepts of systems theory such as hierarchical structures, interactions and holism and control	-It was intendedly developed based on systems theory. -It includes concepts like control, hierarchy, emergency and interactions	-It is based on resilience engineering. -It includes some principles of systems theory such as emergence and interactions
View of accidents	-Accidents are the result of the collective outcome of individuals expressing their degree of freedom while adapting to local constraints	-Accidents occur when external disturbances, component failures, or dysfunctional interactions violate safety constraints (i.e. inadequate enforcement of safety constraints)	-Accidents happen when the variability of the adjustments of everyday performance aggregates in unexpected ways and experiences functional resonance
Basic principles	-Sociotechnical hierarchical structure (mainly for operations) -Flow of information -Boundaries -Gradients and Brownian movements -Safety space of performance	-Sociotechnical hierarchical structure (for development and operations) -Constraints -Emergent properties -Control loops, feedback and process models	-Equivalence of failures and success -Approximate adjustments and variability of performance -Emergence -Functional resonance
Modeling approach	-It models the decisions, actions and information flows across six levels of the sociotechnical structure	-It models the controllers, control mechanisms and feedback loops across the multiple levels of the sociotechnical system	-It models the functions of a system and the dependencies among the functions.
Interesting features	-It can be used to model and analyze the multiple levels of the road transport system and to identify relationships and boundaries of safe performance -The method AccciMap can be used for accident analysis and proactive risk management.	-It explicitly considers automation -It can be used to model and analyze the multiple levels of the road transport system and to identify constraints -Two separate methods for hazard analysis (STPA) and for accident analysis (CAST)	-It can be used to model the functions of the road transport system and its variability of performance -It puts an emphasis on identifying what is needed for everyday performance to go right -It can be used for risk assessment for accident analysis

2.5 STAMP, STPA and CAST as the conceptual framework for the thesis

This section presents the reasons for choosing STAMP as the conceptual framework for the thesis. Next, it provides the background of STAMP, and a detailed description of STAMP, STPA and CAST.

2.5.1 Why STAMP

STAMP was selected as the conceptual framework for this thesis due to four main reasons. Firstly, STAMP has the strongest conceptual connection with key aspects of systems theory such as system structure, the relationships and interactions between components and the system behavior (Underwood and Waterson 2013). This is expected considering that STAMP was deliberately developed to be based on systems theory.

Secondly, STAMP was designed for software and automation; in fact Leveson's earlier work on software safety (Leveson 1995) led her to develop a model that attempted to better understand software-intensive-systems. Consequently STAMP explicitly takes into account some specificities of automation that the other models do not consider. For example, STAMP represents controllers by including sensors, actuators, a process model and a control algorithm. Additionally, the control flaws and hazard causes captured by STAMP are particularly useful for automation e.g. inadequate process model and flawed software requirements.

Thirdly, STAMP provides a functional representation of all the levels of the socio-technical system which can reflect the technical, human and organizational factors within the same frame. While Rasmussen's six-level control structure focuses on operations, Leveson extended the scope to consider both operations and design (Leveson 2017b).

Fourthly, the hazard analysis method and accident analysis method based on STAMP could provide assistance for the application of STAMP to the three research questions. For instance, STAMP and STPA could help to identify hazards related to automated driving systems and automated driving trials in order to address the research questions 1 and 2. Moreover, STAMP and CAST could be applied to address the analysis of crashes involving automated driving.

2.5.2 Background

Accident models are a simplified representation of an accident that provide a framework to help understating how and why accidents occur. They are the basis for accident analysis and investigation methods, hazard analysis methods and accident prevention (Leveson 2011). Leveson argues that the changes of today's systems are stretching the limits of traditional accident models and techniques. Some of these changes include the fast pace of technological change which challenges engineering methods, the changing nature of accidents (mainly introduced by software and digital technology), the new types of hazards, the increasing complexity among system components, the complex interactions between humans and automation. The need for new models capable of dealing with today's systems motivated Leveson to create STAMP.

Traditional accident models

Traditional accident models describe accidents as a result of a sequence of events, each event related to the event that precedes it in the sequence. Moreover, the relationship between cause and consequence is linear and well defined. An example of these models is the Domino model proposed by (Heinrich 1931). In this model, accidents are seen as one of five accident factors that are lined up sequentially like dominoes. When one of the dominoes –accident factors— falls down, it has a knockdown effect that results in an accident which may lead to an injury. As displayed in figure 13, there are five accident factors:

- Ancestry and social environment: undesirable traits of character such as recklessness, stubbornness and greed that may be passed along through inheritance. The environment can also contribute to develop undesirable traits of character or may interfere with education. Both ancestry and social environment may cause faults of person.
- Fault of person: Reasons for committing unsafe acts or for the existence of mechanical or physical hazards.
- Unsafe acts and/or mechanical or physical hazard: Unsafe performance of persons liker errors.
- Accidents: They are caused by unsafe acts/conditions and may lead to injuries.
- Injury: Injuries are the consequences of accidents.

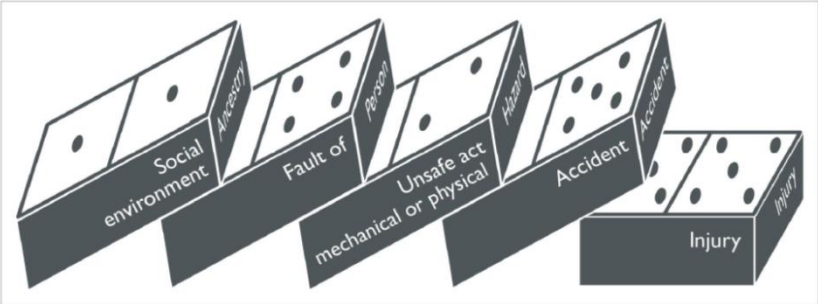


Figure 13 – Domino Model

In this model, each factor is dependent on the previous factors and thus accidents can be prevented by removing one of the preceding factors and interrupting the knock-down effect. Heinrich proposed that unsafe acts and mechanical hazards were the key factor in accidents; removing this factor made the preceding factors ineffective. Although there have been updated versions of the domino model for instance the Loss Causation Model introduced by (Bird and Germain 1996) who introduced the influence of management error, the model is still a chain-of-event representation of accidents.

Another traditional accident model according to Leveson is the Reason's Swiss cheese model (displayed in figure 14). The Swiss cheese model was first introduced as an analogy with the spreading of a disease in which latent failures represent the resident pathogens within the human body, which combine with external factors (stress, toxic agencies, etc) to bring about diseases (Reason 1990). This metaphor was then extended into an organizational accident model, which is commonly known as the Swiss Cheese Model—even if the Swiss cheese variation of Reason's model was not Reason's initiative but a model developed by some of its users and advocated (Le Coze 2013). For this model, Reason proposes to think about the basic elements of a production system: decision makers, line management, preconditions, productive activities and defenses, where active failures and latent conditions can create holes and allow trajectory of accident opportunity.

Active failures are defined as unsafe acts committed by people who are in the sharp end of the organization. They have a direct and usually short-lived impact on the integrity of defenses. Latent conditions are the inevitable "resident pathogens" in the system. They arise from decisions of people who are at the blunt end of the organization. They can translate into provoking conditions within the workplace (time pressures, inadequate equipment,

fatigue, inexperience, etc.) and they can create long-lasting holes or weaknesses in the defenses that may lie dormant within the system. Based on these concepts, Reason defined organizational accidents as situations in which latent conditions –arising from management decisions practices or cultural influences) combine adversely with active errors committed by individuals or terms at the sharp end of an organization, to produce an accident (Reason 1997).

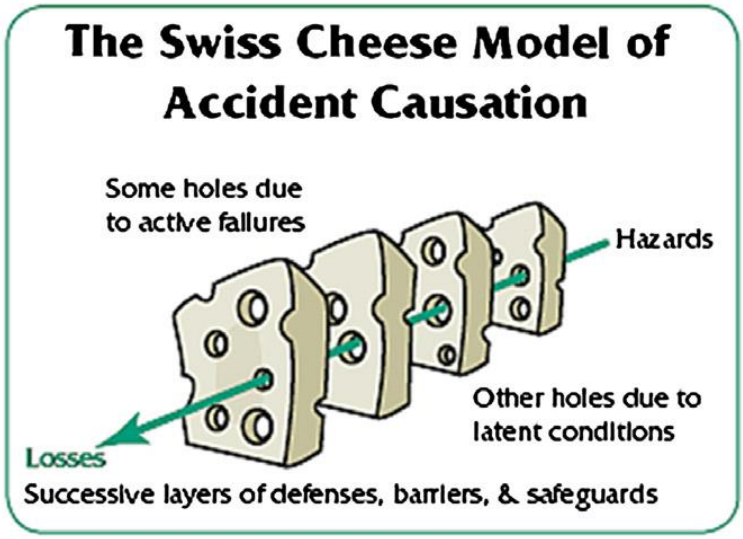


Figure 14 – Swiss Cheese Model (Reason 2008)

Traditional accident models based on chain of failures and events provide the basis for most of the existing accident analysis and hazard analysis methods (Fault Trees, Event Tress, HAZOP, FMEA, etc.) and support the idea that accidents can be prevented by increasing reliability and preventing failures.

In electromechanical systems with simple interactions, safety and reliability have a very close relationship; all the design errors of such systems can be identified during development and testing. However, in complex systems (notably those involving automation) their design cannot be extensively tested and accidents often result from unsafe interactions among components in which there is no failure. Consequently, Leveson suggests that a new accident model based on systems theory rather than reliability theory has to be created.

Furthermore, table 4 illustrates the old assumptions that need to be updated with new assumptions in order to build the model.

Table 4 – Old assumptions and new assumptions (Leveson 2011)

Old Assumption	New Assumption
Safety is increased by increasing system or component reliability; if components do not fail, then accidents will not occur.	High reliability is neither necessary nor sufficient for safety.
Accidents are caused by chains of directly related events. We can understand accidents and assess risk by looking at the chains of events leading to the loss.	Accidents are complex processes involving the entire socio-technical system. Traditional event-chain models cannot describe this process adequately.
Probabilistic risk analysis based on event chains is the best way to assess and communicate safety and risk information.	Risk and safety may be best understood and communicated in ways other than probabilistic risk analysis.
Most accidents are caused by operator error. Rewarding safe behavior and punishing unsafe behavior will eliminate or reduce accidents significantly.	Operator error is a product of the environment in which it occurs. To reduce operator "error" we must change the environment in which the operator works.
Highly reliable software is safe.	Highly reliable software is not necessarily safe. Increasing software reliability will have only minimal impact on safety.
Major accidents occur from the chance simultaneous occurrence of random events.	Systems will tend to migrate toward states of higher risk. Such migration is predictable and can be prevented by appropriate system design or detected during operations using leading indicators of increasing risk.
Assigning blame is necessary to learn from and prevent accidents or incidents.	Blame is the enemy of safety. Focus should be on understanding how the system behavior as a whole contributed to the loss and not on who or what to blame for it.

2.5.3 System Theoretic Accident Model and Processes (STAMP)

STAMP is an accident model based on systems theory in which rather than treating safety as a reliability and failure problem, safety is treated as a control problem managed by a control structure embedded in a sociotechnical system (Leveson 2004, 2011). Further, safety is considered an emergent property that arises from the interactions of system components. Accordingly, STAMP views accidents as a loss of control that arises when external disturbances, component failures, or dysfunctional interactions among system components, violate the safety constraints that the sociotechnical control system imposes on the system behavior to enforce safety. This model captures accidents due to component failure, component interactions, system design errors, human decision making, inadequate controls, flawed safety culture, and flawed organizational design. In STAMP, understanding an accident requires determining why the control structure was ineffective and preventing future accidents requires designing a control structure that will impose the necessary constraint on system behavior to enforce safety. As a result, STAMP changes the emphasis from prevent failures to enforce safety constraints on system behavior.

STAMP has three basic concepts:

1. Safety constraints.
2. Hierarchical control structure.
3. Process models.

1. Safety Constraints

Safety constraints are imposed on the lower levels of the system to control system behavior and enforce safety. The controls to enforce safety constraints can be very broad; they can be related to the physical, human and social components of a system. For instance, in the transport system, the high level control constraint in which vehicles must not violate a minimal safety distance to other road users or objects, is enforced through several controls. An example of physical control involves the way drivers maintain a safe distance by executing control actions on the brakes of the vehicle. Additionally, vehicle systems such as collision avoidance systems can detect imminent collisions and either warn the driver or autonomously execute control actions on the brakes. An organizational control is the standards created by the company management of the vehicle manufacturer or supplier that designs the collision avoidance system. Organizational controls also take place in the operation of the system, for example the procedures of road infrastructure maintenance set by road infrastructure managers. Social controls are provided by stakeholders such as the government who oversees the activities of automakers and road infrastructure companies and defines regulations on road maintenance and the certification of vehicle systems.

2. The hierarchical safety control structure

In STAMP, complex systems are modelled as hierarchical safety control structures with multiple levels where the controllers contained in each level impose safety constraints on the behavior of the controllers and components at the level below. Figure 15 displays a generic hierarchical safety control structure with two basic control structures (one for the system development on the left and another one for the system operations on the right). The figure shows the interactions between the different levels of the structure in terms of control actions issued from higher levels to lower levels (downward arrows) and feedback about the controlled process and the effects of safety constraints on lower controllers,

provided from lower to higher levels (upwards arrows). An example involving the highest levels of the two structures, are controllers such as the congress and legislatures, which impose safety constraints on the levels below via legislation and receive feedback through reports, lobbying, hearings and open meetings, and accidents.

At the company management level, controllers impose safety policies, standards and resources on the lower levels of the company, to design and develop a system, or to operate the system. Moreover, the interactions amongst the two structures can be observed on the communication channels between the companies that develop the systems and the companies that operate the systems. In fact, manufacturers must inform their customers on the assumptions about the operational environment including maintenance and quality procedures and operational procedures. In turn, the operating process provides feedback to manufacturers on the performance of the system, problems and incidents. Lastly, the operating process which is represented by the lowest level of the system, it includes the human operator, automation and the physical process.

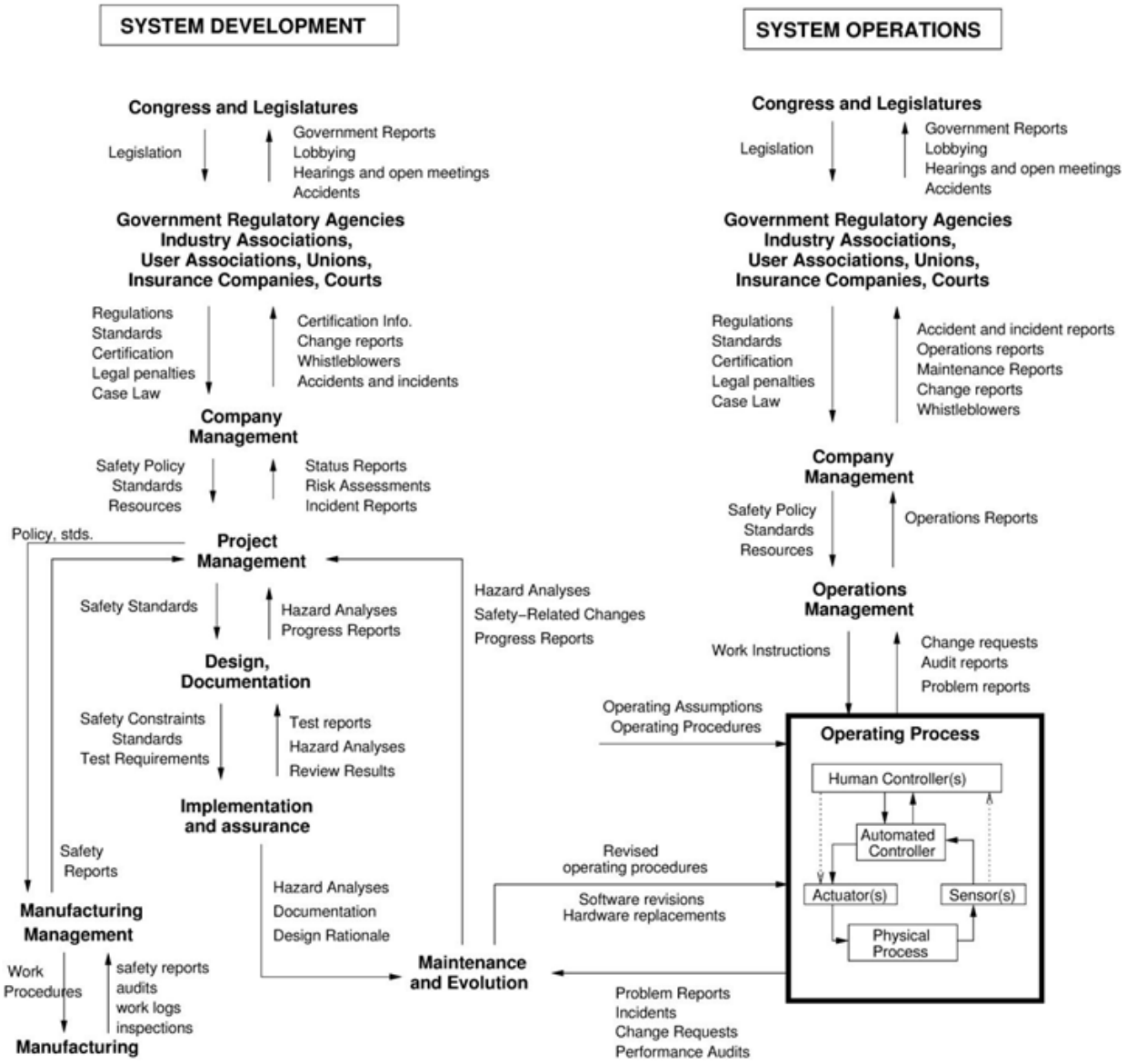


Figure 15 – General model of a sociotechnical system (Leveson 2011)

3. Process Models

As displayed in figure 16, each controller contains a process model (also called mental model for human controllers) that includes the controller's understanding of:

1. The current state of the controlled process.
2. The desired state of the controlled process and the constraints assigned to enforce safety.
3. The ways the process can change. Process models are used by controller's control algorithm to decide what control actions are needed in order to enforce safety

constraints. Additionally, the process models are updated through feedback on the controlled process.

Accidents often occur when the controller's process model becomes inconsistent with the actual state of the process, leading the controller to provide an unsafe control action or to not provide a necessary control action. Also, because process models are kept to date through feedback, accidents also occur when feedback is inappropriate incorrect, missing, or delayed.

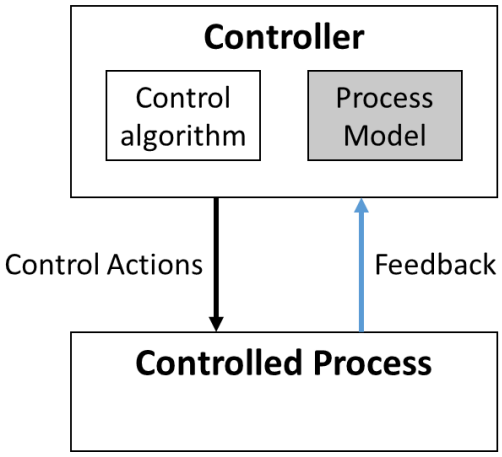


Figure 16 – Process Model

Classification of types of unsafe control actions

STAMP defines the following four types of unsafe control actions (Leveson 2004, 2011; Leveson and Thomas 2013):

1. An unsafe control action is provided that leads to a hazard (e.g. an air traffic controller issues an advisory that leads to loss of separation that would not otherwise have occurred).
2. Not providing a necessary control action leads to a hazard (e.g. the air traffic controller does not issue an advisory required to maintain safe separation).
3. A control action provided with wrong timing (early, late) or in the wrong order leads to a hazard.
4. A continuous control action provided too long or too short a time leads to a hazard (e.g. the pilot executes a required ascent maneuver but continues it past the assigned flight level).

STAMP-based methods

Using STAMP causality model as a foundation, Leveson developed two methods:

- CAST: an accident analysis method to describe and understand an accident that has already occurred.
- STPA: a hazard analysis method to identify potential causes of future accidents.

While other methods based on STAMP have been developed such as STECA, STPA-SEC, Leading indicators, etc., the research in this thesis focuses on the application of STPA and CAST.

2.5.4 STPA

STPA is a hazard analysis method with STAMP as its conceptual foundation; thus it is based on control and systems theory. The goal of STPA is to accumulate information about how hazards can occur (scenarios); this information can then be used to eliminate, reduce, and control hazards in system design, development, manufacturing and operations. It is not designed to develop probability numbers related to the hazards. While traditional hazard analysis techniques were designed to prevent component failure accidents (accidents caused by one or more components that fail), STPA was designed to also address increasingly common component interaction accidents, which can result from design flaws or unsafe interactions among non-failing (operation) components. It identifies more causal factors and hazardous scenarios, particularly those related to software, system design and human behavior.

As seen in figure 17, the STPA process can be divided into four parts; in this sub-section, each part of the process is explained and illustrated with an example involving an STPA analysis on an Adaptive Cruise Control (ACC) system conducted by (Van Eikema Hommes 2012).

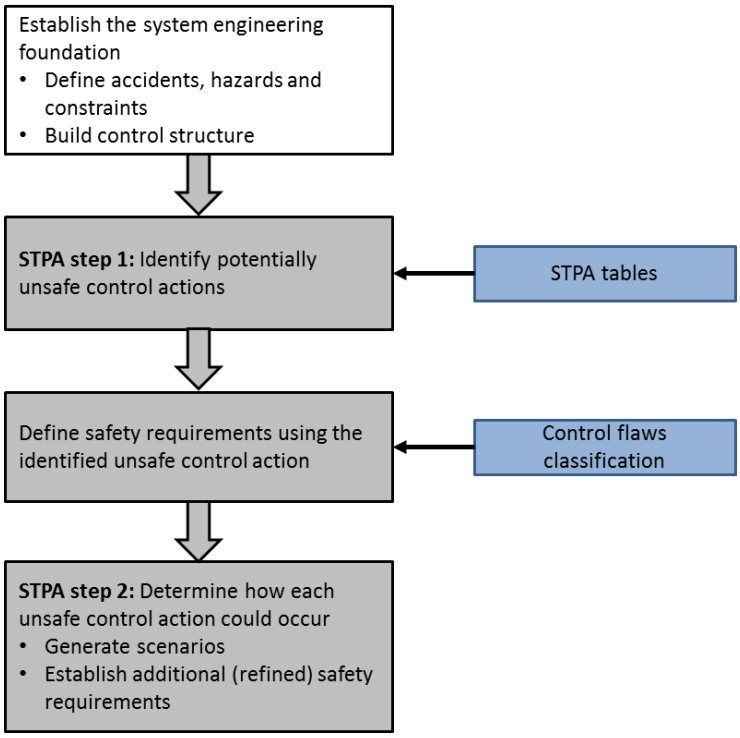


Figure 17 – STPA Process as described in (Leveson and Thomas 2013)

1. Establish the system engineering foundation for the analysis:

Before actually starting the STPA analysis, the system engineering foundation has to be established by defining the accidents, hazards and safety constrains at the system level and by building the control structure.

In STAMP, accidents and hazards are defined as:

- **Accident/loss:** An accident is an undesired and unplanned event that results in a loss, including a loss of human life or human injury, property damage, environmental pollution, mission loss, financial loss.
- **Hazard:** A system state or set of conditions that together with a worst-case set of environmental conditions, will lead to an accident (loss).

Two examples of the accident, hazards and safety constraints for an ACC system are shown in table 5. The first set of definitions taken from (Leveson and Thomas 2013) is larger than the second set taken from (Van Eikema Hommes 2012). In fact, the two hazards and related constraints of the second example could be considered to be a subset of the only hazard and related constraint of the first example. Although there can be some differences in the level of description, the hazards should be associated to high-level system states and thus the number of hazards should be small (less than 10 according to Leveson).

Table 5 – Examples of accidents, hazards and related safety constraints for an ACC system (Van Eikema Hommes 2012; Leveson and Thomas 2013)

Accidents	Hazards	Constraints (Requirements)
Vehicle collision while ACC is engaged	Inadequate distance between vehicle and object in front or in back	Vehicles must never violate minimum separation distance to object in front or in back
Vehicle collision while ACC is engaged	H1: ACC does not maintain a safe distance from the object in front, resulting in collision H2: ACC slows down too abruptly, and vehicle is rear-ended	SC1: ACC must maintain a safe distance from the object in front, resulting in collision SC2: ACC must not brake too abruptly

The next step is to draw the control structure of the system being analyzed; the structure can start as a very simple functional structure covering the controls and responsibilities of the different components, and then it can be refined by adding the control actions and feedback. An example of a basic functional structure and a detailed structure for an ACC system is provided in figure 18.

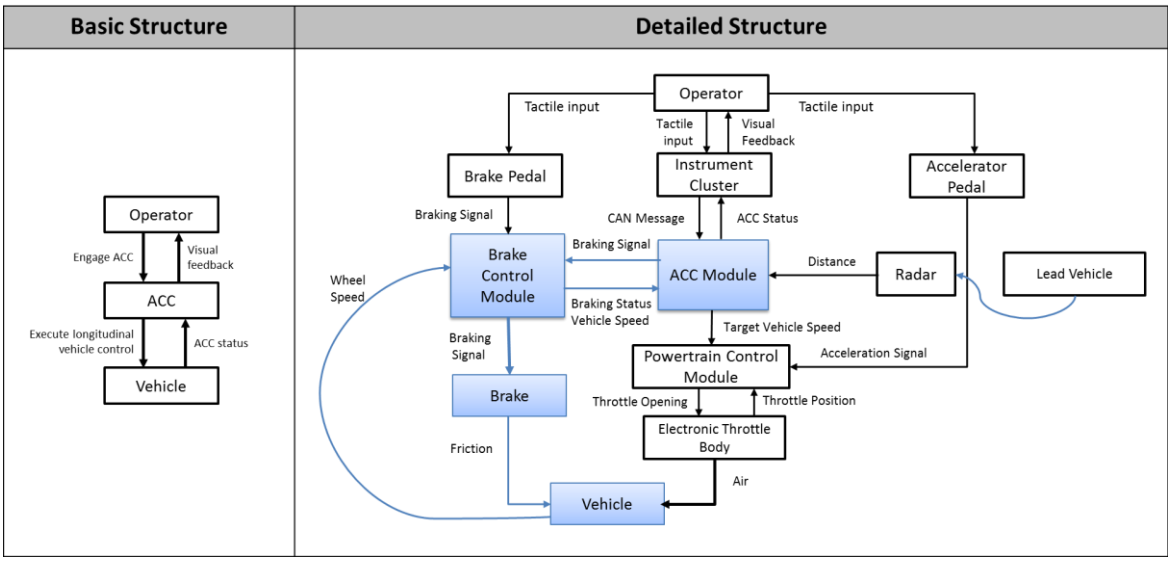


Figure 18 – Basic structure and detailed structure for an ACC system (Van Eikema Hommes 2012)

2. STPA step 1 (Identify unsafe control actions):

The first step of the STPA analysis is to identify the potentially unsafe control actions by examining every control action relative to the four types of unsafe control actions. Additionally, Leveson proposes to document the results of the STPA step 1 using a table as observed in table 6 which contains some unsafe control actions identified for the ACC system (Van Eikema Hommes 2012).

Table 6 – Examples of unsafe control actions for an ACC system

Hazard: ACC does not maintain a safe distance to object in the front				
Control Action	**Not providing CA causes hazard**	**Providing CA causes hazard**	**Wrong timing/order of CA causes hazard**	**CA stopped too soon/applied too soon**
Brake signal from ACC to Brake Control Module	Vehicle does not brake when the distance to lead vehicle is less than the value set by the operator	Commanded deceleration is too low when the distance to lead vehicle is less than the value set by the operator	Braking is commanded too late when the distance to the lead vehicle is less than the value set by the operator	Braking stops before safety distance between the vehicles is reached

3. Define safety requirements and constraints:

Once the unsafe control actions have been identified, they are used to define safety requirements and constraints[6]. To this end, the unsafe control actions are translated into safety requirement and constraints as illustrated below:

- **Unsafe control action:** Vehicle does not brake when the distance to lead vehicle is less than the value set by the operator.
- **Safety requirement:** Vehicle must brake when the distance to lead vehicle is less than the value set by the operator.

4. STPA step 2:

The potential causes of (scenarios leading to) unsafe control actions being provided and of required safe control actions not being executed, are generated and used to define additional safety requirements; I refer to these requirements as refined safety requirements (as you will see in chapters 3-5) to distinguish them from the safety requirements defined based on the unsafe control actions.

Determining how potential hazardous control actions could occur involves examining the component and interactions within the control structure; it requires prior experience with the system and creativity in order to come up with valid and plausible scenarios. Further, Leveson provides a classification control flaws related to the control loop (displayed in figure 19) to assist the elaboration of scenarios.

[6] Constraint is the term used in systems theory and requirement is a term more commonly used in engineering.

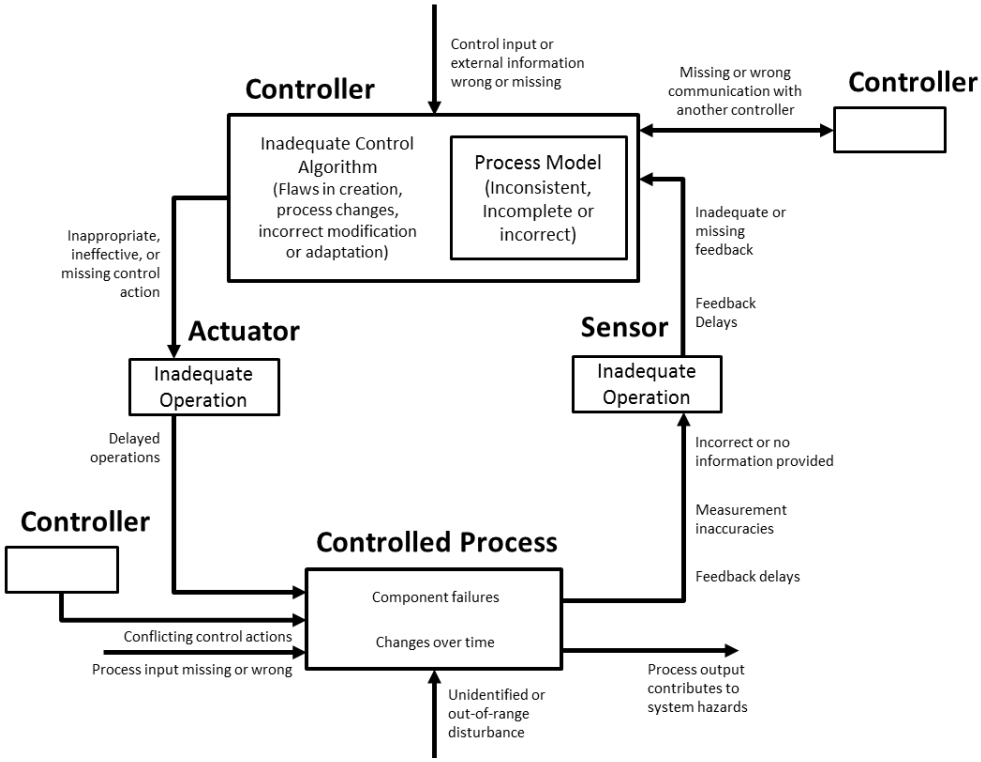


Figure 19 – Potential control flaws related to the control loop

Table 7 displays an example of the analysis for the unsafe control action in which the vehicle does not brake when the distance to the lead vehicle is less than the value set by the operator. It shows a scenario associated to the control flaws identified in the feedback loop that provides information to the sensors of the brake control module.

Table 7 – Examples of scenarios for the ACC system (Van Eikema Hommes 2012)

UCA: Vehicle does not brake when the distance to lead vehicle is less than the value set by the operator		
Scenario	**Associated causal factors**	**Refined safety requirement**
The brake control module (BCM) receives incorrect or no information or delayed information regarding wheel rotation signals.	Accumulation of dirt on wheel rotation sensor	The BCM must detect when wheel rotation signal is inaccurate
	Wire disconnection between brake control module and sensors	The BCM must detect when there is no wheel rotation signal
	Communication bus fault such as message priority	The communication bus must include a prioritization of messages which ensures that messages reach the brake control module in time

The information about hazards accumulated through the STPA analysis (i.e. safety requirements, scenarios and refined safety requirements) can then be used to eliminate, reduce, and control hazards in system design, development and operations (Leveson and Thomas 2013).

2.5.5 CAST

CAST is an accident analysis method with the STAMP model as its foundation. Accordingly, it assumes that accidents are caused by an inadequate enforcement of safety constraints and extends the view of accident causation from component failure accidents to include other causes like design errors, unintended and unplanned interactions among system components, flawed safety culture and human decision making, inadequate control and oversight, and flawed organizational design.

The goals of CAST are (Leveson 2017a):

- To Provide a framework and process to assist in understanding the entire accident process and identifying the systemic factors.
- To get away from blame (who) and shift the focus to (what) and how to prevent such occurrences in the future.
- Identify why people behaved the way they did, including the contextual factors that influenced their behavior.
- Minimize hindsight bias.
- Determine the weaknesses in the safety control structure that allowed the loss to occur.

As displayed in figure 20, the CAST process can be divided into five parts; like for the STPA method, this sub-section explains and illustrates each part of the process using an example. The example comes from a CAST analysis on a chemical industry accident involving an explosion and fire at the Shell Moerdijk plant. This analysis was recently created by Leveson as a benchmarking exercise (Leveson 2017a).

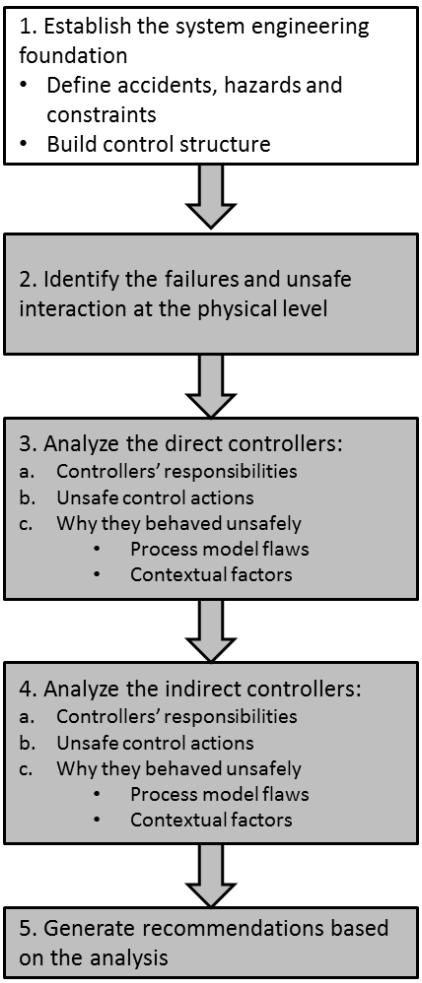


Figure 20 – CAST process (Leveson 2011, 2017a)

1. Establish the system engineering foundation:

A CAST analysis starts with the identification of the hazards that led to the accident (or loss) and the safety constraints that must be satisfied to prevent the loss. The hazards and constraints of the explosion and fire that occurred at the Shell Moerdijk plant The Netherlands on June 3 2014 are displayed on table 8.

Table 8 – Examples of hazards and constraints for the Moerdijk accident (Leveson 2017a)

Hazard	Safety Constraints
H1: Exposure of public or workers to toxic chemicals	**1.** Workers and the public must not be exposed to potentially harmful chemicals **2.** Measures must be taken to reduce exposure if it occurs
H2: Explosion (uncontrolled release of energy and/or fire	**1.** Chemicals must be under positive control at all times **2.** Warnings and other measures must be available to protect workers in the plant and minimize losses to the outside community **3.** Means must be available, effective, ad used to respond to explosions or fires in the plan

After the definition of the hazards and constraints involved in an accident, the control structure at the time of the accident is modeled. The CAST analysis examines each component of the control structure and their interactions to understand how they contributed to the loss. The basic control structure and detailed control structure for Shell Moerdijk is displayed in figure 21.

2. Identify failures and unsafe interactions at the physical level:

The second step consists of analyzing the physical level of the system to identify failures and unsafe interactions that contributed to the accident. In the Moerdijk accident there were no physical control fails; conversely, there were unexpected and unsafe chemical and physical interactions which caused the physical collapse of the reactor and separation vessel. For example, the process to distribute the ethylbenzene over the catalyst pellets (wet them) resulted in dry zones and gas formation increased the pressure in the reactor (Leveson 2017a).

3. Analyze direct controllers:

The analysis of controllers starts with the controllers immediately above the physical process (i.e. direct controllers). The analysis comprises the identification of:

- The safety constraints and responsibilities enforced by controllers to prevent the loss.
- The unsafe control actions.
- The reasons why controllers behaved unsafely by examining the process model flaws and contextual factors.

Some examples of the analysis for the Process Control System and the human operators in the Moerdijk accident are summarized in table 9.

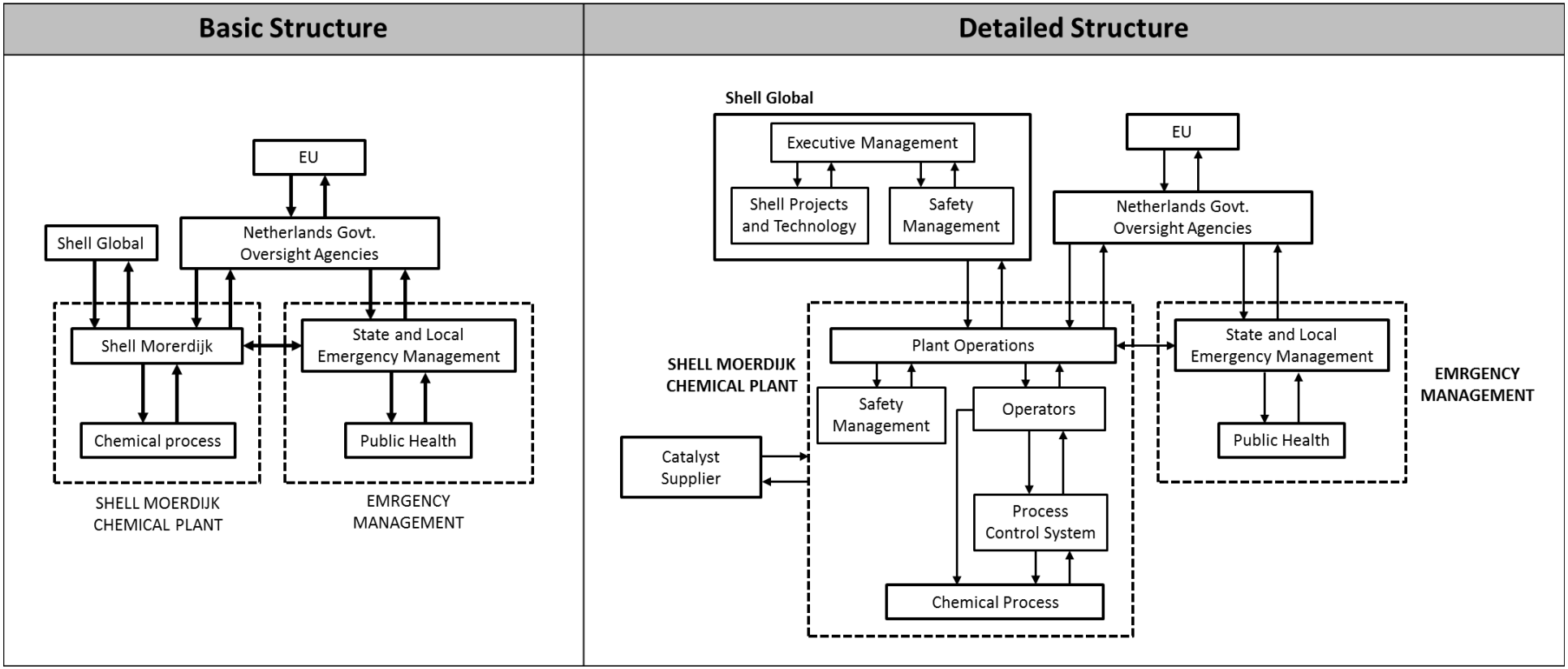


Figure 21 – Basic Control Structure and Detailed Control Structure for Shell Moerdijk

Table 9 – Excerpt of the direct controllers' analysis for the Moerdijk accident

Analysis	Process Control System	Human Operator
Responsibilities	-Assist operators in controlling the plant during normal production and off-nominal operations -Display relevant values, provide, issue control actions on plat equipment -Control temperature, pressure, level, and flow to ensure that the process remains within the safe margins and does not end up in an alarm situation	General: -Monitor plant conditions and alarms -Control the process such that it stays within safe boundaries of operation -Respond to unsafe conditions that occur Specific: -Adjust gas and liquid flows as needed during startup -Make sure the Unit is not heated too quickly
Unsafe control action	The process control system did not provide the assistance required by the operators to safely control the start-up process including automatically controlling the heating rate and other important variables	The operators did not stabilize or halt the process before the explosion when critical process boundaries were executed
Model flaws	The process control system had the correct information to help operators but was not configured to provide the necessary help to the operators during a start-up	The operators were not aware that the situation was dangerous; they did not know that critical conditions had been exceeded and therefore did not decide to intervene.
Contextual factors	NA	-The panel operator and production team were experienced staff on Unit 4800 but had never experienced a startup of Unit 4800 after a catalyst change -The controlled process system was configured for production and not for start-up -Work instructions were incomplete

4. Analyze indirect controllers:

The analysis then moves upward in the control structure in order to examine the indirect controllers of the system. As for direct controllers, the analysis of indirect controllers involves the identification of: the safety constraints and responsibilities enforced by controllers to prevent the loss, the unsafe control actions and the reasons why controllers behaved unsafely by examining the process model flaws and contextual factors.

Some examples of the analysis of indirect controllers in the Moerfijk accident are illustrated in table 10.

Table 10 – Excerpt of the Dutch regulators' analysis for the Moerdijk accident

Analysis	Dutch regulators
Responsibilities	General -Supervise and enforce Dutch laws to protect the environment and the public -Enforce the EU health and safety laws within the Netherlands Specific -Identify shortcomings at companies they are responsible to oversee -Encourage companies to improve their safety-critical process through supervision and enforcement -Assess modifications mad to plants, processes, and processes -Pay greatest attention to safety-critical process
Unsafe control action	Did not identify shortcomings at Shell. Assessed Shell as a well-functioning company in which they had a great deal of confidence.
Model flaws	Regulators had a positive view of the Shell Moerdijik safety management system.
Contextual factors	-The regulatory agencies had scarce resources and time for oversight -Shell had only one violation between 2010 and 2014, and always initiated improvement actions when a problem was identified -Several shortcomings at Shel Moerdijk were not labeled as violations

5. Generate recommendations:

The last step of the process is to use the results of the analysis to establish recommendations that will eliminate or reduce unsafe behavior.

Some examples of the recommendations generated in the CAST analyses are (Leveson 2017a):

- The physical design limitations and inadequate physical controls need to be fixed.
- The process control system should be redesigned to assist operators in all safety-critical, off-nominal operations.
- A human factors study during the job is needed to ensure that operators are provided with information and a work situation that allows them to make appropriate decisions, better automated assistance should be provided in all phases of operation, training should be provided for activities that are known to be hazardous, and work instructions as well as the process for producing them need to be improved.
- Dutch regulators: Better supervision of the highest risk activities is needed; they need to oversee and ensure that strict procedures are being used for the most dangerous activities and that safety management system is operating effectively and following their own rules.

2.5.6 The STAMP-based approach of the thesis

Chapters 3-5 of the thesis use STAMP, STPA and CAST to examine the safety benefit assessment, trial safety and accident analysis of automated driving. In chapter 3, STPA is used to conduct an analysis of the human driver and automated controllers of a highway pilot system. In chapter 4, STPA is used to conduct two analyses; a first analysis on the vehicle trial process (which includes high-level controllers) and a second analysis on a vehicle trial operation involving a highway pilot system. Lastly, in chapter 5, STAMP concepts are used to establish guidance elements for the analysis of automated driving systems, and CAST is employed as the backbone of a new accident analysis method for crashes involving automated driving.

Résumé chapitre 3: Evaluation des gains de sécurité

Afin d'apporter une réponse à la première question de recherche « le véhicule autonome améliore-t-il la sécurité routière ? » le chapitre 3 contribue à l'évaluation des gains de sécurité attendus avec le système de conduite autonome. Le système considéré dans ce chapitre est celui de conduite autonome sur autoroute (Highway Pilot System), qui sera certainement l'une des premières applications du véhicule autonome à être déployée.

Faire cette évaluation requiert dans un premier temps de déterminer la population cible du système et de calculer son efficacité afin de fournir des estimations quantitatives en termes de réduction d'accidents et de diminution des dégâts corporels.

Tandis que la population cible peut être estimée à partir des bases de données d'accidents existantes, le calcul de l'efficacité d'un nouveau système comme celui de la conduite autonome implique de mener une analyse prospective. Cette analyse repose sur un cadre conceptuel de neuf mécanismes de sécurité couvrant les trois dimensions de la sécurité routière (risque, exposition et conséquence) et sur des données empiriques (études, expérimentations, enquêtes,..) pour quantifier les effets des mécanismes. L'évaluation des mécanismes de sécurité nécessite de définir des questions permettant de la cadrer.

Ainsi, la première partie du chapitre donne l'estimation de la population cible du système de conduite autonome considéré et la seconde partie s'attache à définir les questions nécessaires à l'évaluation des mécanismes de sécurité au moyen d'une analyse STPA. Cette analyse structurée a permis d'identifier les contraintes de sécurités dont découlent les questions.

Chapter 3: Examining the safety benefit assessment of automated driving systems

3.1 Chapter overview

As displayed in figure 22, the safety benefit assessment of vehicle systems is a quite large process, which requires determining the crash target population that could be potentially addressed by the system and evaluating the system's effectiveness in order to provide quantitative estimates in terms of crash and injury reduction. While the target population is estimated by querying crash databases, the effectiveness assessment of new systems such as automated driving systems (ADS) involves conducting a prospective analysis. Prospective analyses consider a conceptual framework of nine safety mechanisms covering the three dimensions of road safety (risk, exposure and accident consequence) and empirical evidence (e.g. exposure data, studies on driving simulators, field operational trials, questionnaires) to quantify the effects of the nine safety mechanisms. Furthermore, multiple assumptions are made in the evaluation of the nine safety mechanisms, for example the system is assumed to have a proper functioning and operation, and drivers are assumed to operate the system as designers expect.

Due to the lack of empirical data regarding the safety mechanisms for automated driving, and to the time and resource constraints associated to a thesis, the research presented in this chapter (illustrated with the orange and blue boxes in figure 22) is limited to a partial contribution to the target population and the assumption evaluation. The contribution to the target population consisted of querying a crash database to estimate the target population of a highway pilot system in terms of crash frequency, fatalities, and injuries, and relative to all crashes. On the other hand, the contribution to the assumption evaluation comprised conducting an STPA analysis on a highway pilot system to identify safety requirements related to the driver and automation, and using those safety requirements to define questions that assist a further examination of the assumptions associated to the evaluation of the direct safety mechanisms. The results of the target population estimates, the usage and outputs of the STPA

analysis and the questions based on safety requirements are discussed. Lastly, new possibilities for this research are described, notably the availability of empiric data from the L3Pilot project which will enable to implement the questions and quantify the safety benefit.

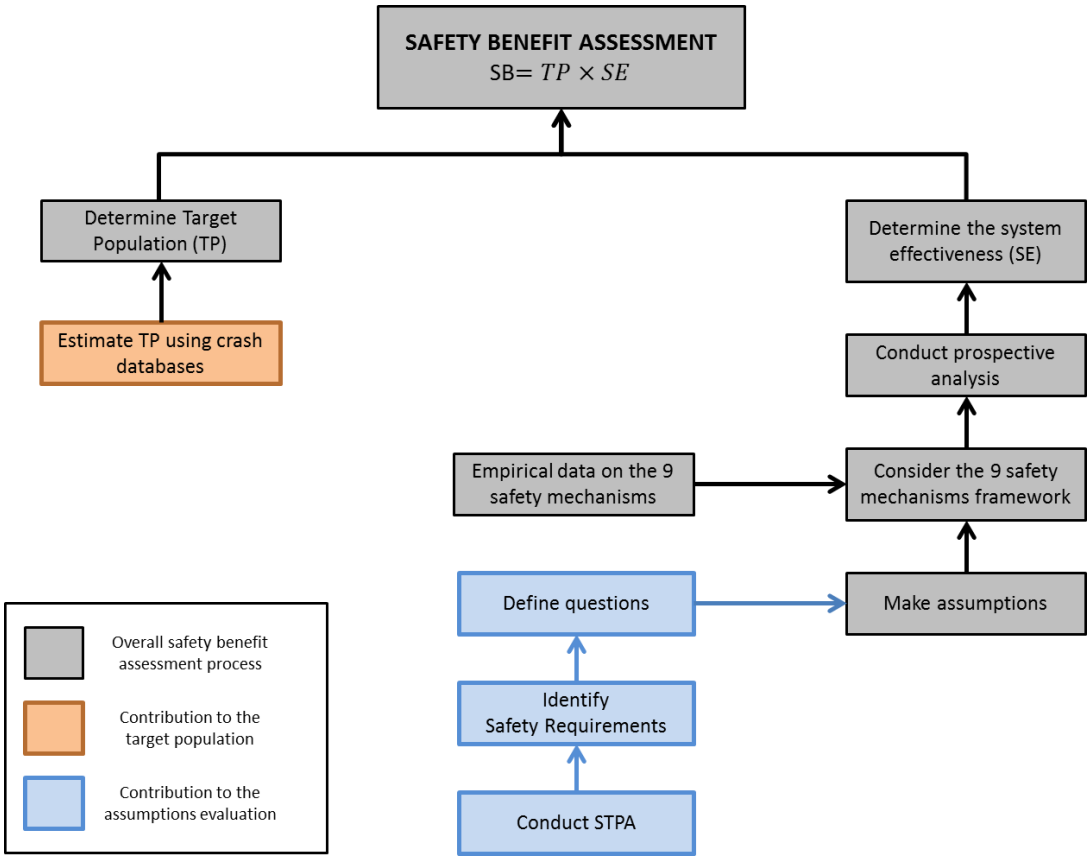


Figure 22 – Overall process of the safety benefit assessment and chapter's contribution

3.2 Introduction

Automated driving systems are expected to improve road safety, nevertheless the road transport stakeholders need to assess the safety benefit of Automated Driving Systems (ADS) to have a better understanding of their safety effect and to support decisions regarding ADS (Risto Kulmala 2010). For instance, the authorities and policy-makers consider the safety benefit of ADS—along with other aspects such as their feasibility, acceptability and efficiency—to decide whether or not to endorse and invest on ADS. Moreover, the automotive industry also takes into account the safety benefit of ADS to support the decision of which systems to develop.

The fundamental equation for the safety benefit assessment is:

$$SB = TP \times SE$$

Where:

SB= Safety benefit obtained from the implementation of a vehicle system;

TP= Addressable crash population (target population); and

SE= Effectiveness of vehicle system

According to the equation, the target population and the effectiveness of a vehicle system must be determined in order to assess its safety benefit. The target population can be determined by identifying the variables that describe the crashes in which the system could mitigate or avoid a crash (e.g. accident type, location, speed, etc.) and querying crash databases to estimate the addressable crash population in terms of crash frequency, fatalities, and injuries. For example, (Rau, Yanagisawa, and Najm 2015) proposed a methodology that uses variables on location, pre-crash scenarios, driving conditions, travel speed and driver condition and two databases: General Estimates System (GES) and Fatality Analysis Reporting System (FARS), to estimate the target crash population of automated driving systems corresponding to SAE levels 2-4.

Effectiveness assessment evaluations are conducted to determine the effectiveness of vehicle systems; retrospective (aka ex post or a posteriori) effectiveness evaluations are performed for existing vehicle systems on which sufficient crash and exposure data is available (Page et al. 2007), and prospective (aka ex ante or a priori) effectiveness evaluations are performed for new vehicle systems or vehicle systems with low-penetration rates on which there is no crash data (Karabatsou et al. 2007). Most of the ADS systems are technologies which are under development or with low-penetration rates, consequently their effectiveness assessment needs to be conducted via prospective analyses that incorporate automated driving data from studies on driving simulators, on test tracks, Field Operational Trials (FOT's), etc.

Two impact assessment schemes have included directions for the prospective safety assessment of ADS. Firstly, (S. Smith et al. 2015) established a scheme for the impact assessment of automated driving systems that proposes to estimate the safety benefit of ADS through the Safety Impact Methodology (SIM) (Carter et al. 2009). The SIM uses a computer-based simulation tool to simulate the vehicle kinematics and driver/vehicle reaction times with

and without the vehicle system in conflict or crash situations, and estimates the safety benefit by comparing the outputs of the simulations in terms of crash prevention. Virtual simulation considers the events up to a few seconds before a crash, and therefore it primarily takes into account the causal factors that are close to the crash; it does not explicitly incorporate the causal factors that play a role in crashes long before the crash happens, such as inadequate infrastructure design and human's overreliance on automation.

Secondly, the scheme established by the Trilateral Working Group on Automation in Road Transportation (ART WG) (Innamaa, Smith, and Uchida 2016) proposes to go beyond the events close to the crash and to consider the framework for the safety assessment of Intelligent Transport Systems defined by (Draskóczy, Carsten, and Kulmala 1998; Risto Kulmala 2010) to conduct the safety assessment of ADS. As seen in table 11, the safety assessment framework consists of nine safety mechanisms that have an influence on the three dimensions of road safety: exposure, risk and consequence; the dark blue color in the table indicates that the mechanism is focused on the safety dimension, and light blue indicates that the safety dimension is relevant to the mechanism.

Table 11 - The safety assessment framework relative to the three safety dimension adapted from (Risto Kulmala 2010)

Safety mechanism	**Safety dimension**		
	Exposure	**Risk**	**Cons.**
1. Direct modification of the driving task			
2. Direct influence by infrastructure			
3. Indirect modification of user behavior			
4. Indirect modification of non-user behavior			
5. Modification of interaction between road users			
6. Modification of exposure			
7. Modification of modal choice			
8. Modification of route choice			
9. Modification of accident consequences			

All of the nine safety mechanisms are useful for a comprehensive assessment of the safety effects of vehicle systems; however, due to the lack of empirical data on indirect mechanisms

(3-5), and on modification mechanisms (6-7), the application of the framework mainly focuses on the evaluation of direct effects (mechanisms 1 and 2) and relies on expert judgement for the indirect mechanisms (3-5) and mechanism 9, and on questionnaires on users' attitudes for mechanisms 6-8 (Kulmala et al. 2007; Silla et al. 2017).

The assessment of the first direct mechanism involves assumptions on the inherent safety of the vehicle system such as: the vehicle system has a proper and reliable functioning, the interactions with other vehicle systems can be overlooked, the performance of the vehicle system is considered stable rather than variable, etc.; the assumptions on the interactions between the driver and the vehicle systems for example, the driver understands and accepts the system, the driver uses the system as designers expect it, there is no misuse, etc. Further, the assessment of the second safety mechanism implies assumptions on the interactions with infrastructure, such as ignoring the effect of degraded infrastructure conditions and considering that the information received through digital infrastructure like networks is always correct.

The complexity and new roles of the human and the vehicle introduced by automated driving systems require a further examination of the assumptions related to direct mechanisms (1-2). For instance, the assumption that the automated driving system has a proper functioning and a safe operation, necessitates the evaluation of the vehicle sensor's capability to provide adequate feedback on obstacle detection, object classification, and the road environment. Also, these assumptions need the assessment of automation's ability to be aware of its operational design domain and to understand and predict other road users' intentions. Another example comprises the assumption that the driver will respond to the takeover request, which necessitates the evaluation of the HMI design and reliability, the coherence and understandability of the information provided by the HMI, and the driver's knowledge and experience needed for takeover validations. A final example involves the assumption that networks provide correct information to automation, which requires assessing the consequences of missing feedback or delayed information.

STPA provides a method for the systematic identification of safety requirements that allow deriving questions to further examine the assumptions associated to the assessment of the direct safety mechanisms (1-2). For example safety requirements on vehicle sensors such as

"the vehicle sensors must take accurate on-time measures on the driving environment", enables to generate a question on whether or not the vehicle sensors take accurate on-time measures, which in turn enables to further investigate the assumption related to the proper functioning of the vehicle system.

Accordingly, two topics should be investigated in order to assist the broader process of safety benefit assessment of automated driving systems:

- The target population of ADS systems.
- The application of STPA analysis to identify safety requirements and corresponding questions that address the assumptions related to the evaluation of the direct mechanisms.

3.2.1 Aim and objectives

The motivation for this chapter was the first research question *"will automated driving improve road safety?"* which entails assessing the safety benefit of automated driving systems. The aim of the chapter is to contribute[7] to the broader process of safety benefit assessment by estimating the target population of an automated driving system and by providing assistance to evaluate the direct safety mechanisms.

Several objectives were established to achieve this:

- Determine the target population of the highway pilot system by querying crash databases.
- Define safety requirements at the operational level by conducting an STPA analysis on the highway pilot system at the microscopic level.
- Generate questions using the safety requirements, to assist the evaluation of assumptions related to direct safety mechanisms (1-2).

[7] Merely a partial contribution to the safety benefit assessment was made because the lack of solid empirical data regarding automated driving and the 9 safety mechanisms, and the time and resource constraints of the thesis prevent the evaluation of all of the 9 safety mechanisms and the assessment of quantitative estimates.

3.3 Methods

This section presents the methods employed to establish the description of the highway pilot system used as case study, to estimate the target population of the highway pilot system, to perform an STPA analysis on the highway pilot system and thus identify safety requirements, and to define questions (based on the safety requirements) to assist the evaluation of the direct safety mechanisms.

3.3.1 Highway pilot system description

Company documents such as design reviews, technical notes, safety principles, customer requirements, milestone presentations, etc., were reviewed to develop a functional description of the highway pilot system. While the description is inspired in an automated driving system that is currently being developed at Renault, it simplifies certain aspects (e.g. other vehicle systems like the cruise control or emergency brake assist are not considered) and does not reflect the final version of the system which is still being modified as a part of the testing and validation process.

3.3.2 Estimation of the target population

The target population of the highway pilot system was estimated as described by (Rau, Yanagisawa, and Najm 2015). The variables that describe the crashes potentially addressed by the highway pilot system were identified and used to query the crash data in the "Bulletin d'Analyse d'Accident Corporel de la Circulation" (BAAC) database for the year 2015 in order to estimate the target population in terms of the crash frequency, fatalities, and injuries. Additionally, the target population estimates were compared relative to the annual frequencies of all crashes.

3.3.3 Identification of the safety requirements through STPA

Safety requirements related to the highway pilot system were identified by conducting an STPA analysis comprising four parts:

1. Definition of the system engineering foundation: the system engineering foundation for the analysis was established by defining the accidents, hazards and constraints at the system level, and by building the control structure of the highway pilot system at the

microscopic level (including the human driver controller and the automated controller). The control structure was built using the highway pilot description to define the control actions and feedback loops among the system components; further, three company employees working in the design of the highway pilot system, reviewed and validated the control structure.

2. Identification of unsafe control actions (STPA step 1): unsafe control actions were identified by filling-out UCA tables in which every control action of the highway pilot system is examined relative to the four types of unsafe control actions. Additionally, a graphic timeline was generated to map the control actions and unsafe control actions to the driving mode phases and transitions.
3. Definition of safety requirements: the identified unsafe control actions were used to establish safety requirements. For instance, an unsafe control action in which automation provides control of the vehicle during manual driving mode, can be translated into the following safety requirement “automation must not provide control of the vehicle during manual driving mode”.
4. Elaboration of scenarios leading to unsafe control actions (STPA step 2) and definition of refined safety requirements: Due to the large amount of unsafe control actions identified in the first step of the STPA, the unsafe control actions were classified into six categories to reduce their number and thus optimize the analyses processing time of the elaboration of scenarios. Subsequently, the scenarios leading to the six categories of unsafe control actions were generated by examining the control flaw classification proposed by Leveson (Leveson and Thomas 2013). Lastly, the generated scenarios were used to define refined safety requirements.

3.3.4 Definition of questions to assist the evaluation of direct safety mechanisms

The first step involved reviewing all the safety requirements and refined safety requirements to assign the requirements to the first direct safety mechanism one, the direct safety mechanism two, or both. The second step consisted of defining questions based on the requirements, to assist the evaluation of the direct safety mechanisms. Finally, the results of the two steps were organized into two tables (one per safety mechanism) which include the category being

analyzed, the specific safety requirements considered, the questions derived from the safety requirements, the specific refined safety requirements considered and the questions derived from refined safety requirements.

3.4 Findings

Section 3.3 covers the highway pilot system description, the results of the target population estimation for the highway pilot system, the outputs of the STPA analysis (which include the safety requirements), and the tables containing questions based on safety requirements that provide assistance in the evaluation of the two direct safety mechanisms (1-2).

3.4.1 Highway pilot system

The automated driving system analyzed in the STPA analysis is a pilot system for highway use during traffic jams or during long-distance trips, that allows drivers to have their feet off the pedals, hands off the steering wheel and eyes off the road while the automated driving system is engaged. Additionally, the automated system has two configurations: (1) the vehicle stays within its lane and can go up to 90 km/h; and (2) the vehicle is capable of changing lanes and can go up to 110 km/h.

To engage the ADS, automation verifies that the ADS mode availability conditions (e.g. ADS compatible road section, vehicle speed, vehicle position within its lane, etc.) are met before sending a notification indicating that the AD mode is available. Next, the driver must validate the ADS engagement by simultaneously pushing two buttons in the Human Machine Interface (HMI). Once the ADS is engaged, automation takes over the dynamic driving task, i.e. the execution of lateral and longitudinal control of the vehicle, the monitoring of the driving environment, object and event detection and response, and performing the dynamic driving task fallback (SAE International 2016). As a result, the driver can release the control of the vehicle and perform secondary activities, such as reading emails and watching a movie.

Moreover, the driver can initiate ADS disengagement whenever s/he wants, by simultaneously pushing two buttons in the HMI or by overriding the system (i.e. executing actions on the pedals or steering wheel). Furthermore, automation can also initiate ADS disengagement when

automation detects that the vehicle will no longer be/is no longer within its operational design domain or when there is performance-relevant system failure; there are three types of end modes for ADS disengagements initiated by automation:

1. **ADS end mode type 1:** Automation determines that the ADS compatible road section is coming to an end via the navigation system (e.g. the vehicle reaches a highway exit) and starts the ADS end mode type 1. As described in figure 23, a notification is sent to the driver before the end of the ADS compatible road, to prepare the driver for the takeover of the control of the vehicle. If the driver does not takeover, automation sends a takeover request a few seconds before the end of the ADS compatible road. Finally, if the driver does not validate the takeover request, automation performs a minimal risk maneuver which involves slowing down the vehicle to a complete standstill within the same lane (for the first configuration of the ADS) or bringing the vehicle to a complete standstill in the emergency lane (for the second configuration of the ADS).

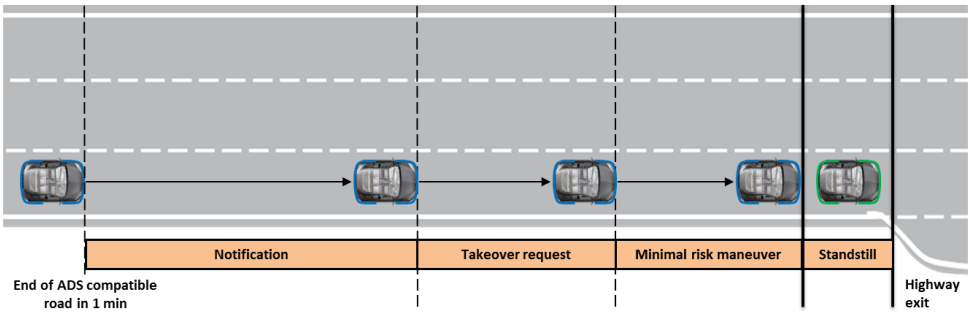


Figure 23 – ADS end mode type 1

2. **ADS end mode type 2:** While the ADS is engaged, automation constantly evaluates the state of the ADS conditions (e.g. heavy traffic, distance to other vehicles, vehicle sensor performance, etc.) to determine if the ADS is within its Operational Design Domain (ODD) and can continue to operate the vehicle. When ADS conditions are no longer met, automation launches the ADS end mode type 2 (illustrated in figure 24) by sending a quick takeover request to the driver. As in ADS end mode type 1, in the absence of the driver's intervention, automation performs a minimal risk maneuver.

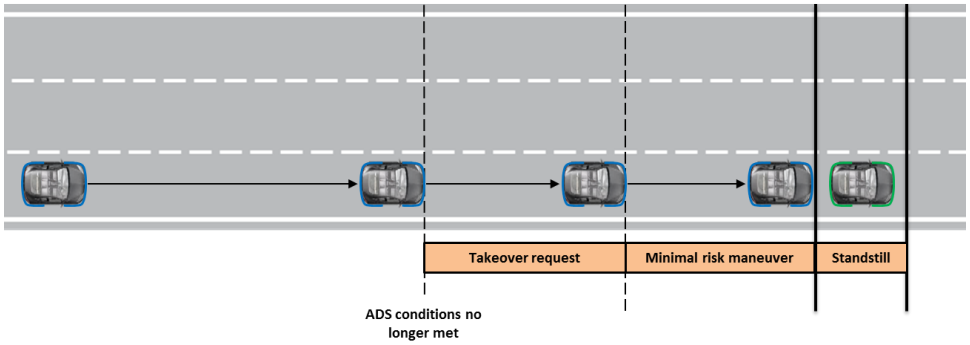


Figure 24 – ADS end mode type 2

3. **ADS end mode type 3:** The third end mode, is launched when there is a performance-relevant failure or when ADS conditions are no longer met, which prevent automation from performing the safe operation of the vehicle (e.g. performance-relevant failure on vehicle sensors that prevent automation from perceiving the driving environment or a broken tie rod). As displayed in figure 25, the ADS end mode type 3 immediately starts the minimal risk maneuver without requesting the driver for a takeover request.

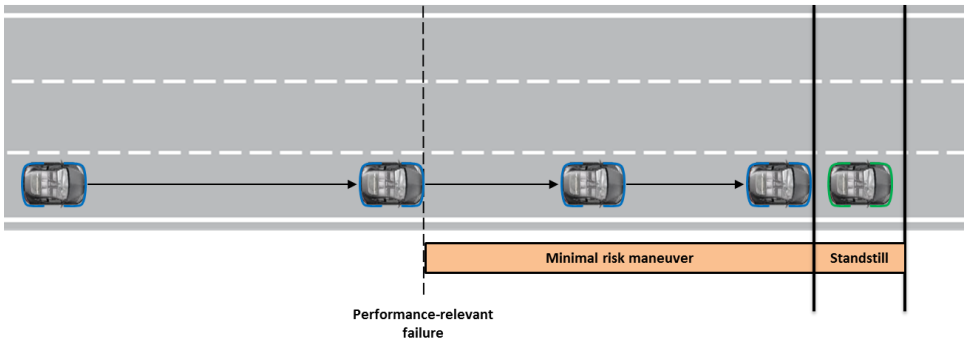


Figure 25 – ADS end mode type 3

3.4.2 Target Population

The crash database used for the estimation of the highway pilot system's target population is described, and then the characteristics and corresponding variables of the crashes which could be potentially addressed by the highway pilot system are listed. Finally, the identified crash variables are used to query the crash database for crashes in 2015, and determine the target population estimates in terms of crash frequency, fatalities, and injuries, and relative to the annual frequency of all crashes.

BAAC

The “Bulletins d’Analyse d’Accident Corporel de la Circulation” (BAAC) are reports established by the French police officers that contain the information they collect on all the “accidents corporels”, i.e. the injury road crashes, in which they intervene. These reports are grouped into a crash accident database with approximately 145 variables describing the crash situation, the location (including infrastructure) of the crash, the vehicles and road users involved in the crashes. The BAAC database is managed by the “l’Observatoire National Interministériel de la Sécurité Routière (ONISR)”and is employed by the government to monitor the state of road safety in France.

The following definitions are useful to understand the scope of the crash information contained in the BAAC database:

- **Injury crash:** a crash on public roads that involves at least one vehicle and causes at least one victim.
- **Victim:** road user involved in a crash who needs medical care.
- **Fatal crash:** an injury crash that causes at least one fatality.
- **Fatality:** Victim killed in the crash or from their injuries up to 30 days after the crash.
- **Injured and hospitalized:** Injured road users who are hospitalized for more than 24 hours.
- **Injured and not hospitalized:** Injured road users who need on-the-spot medical care and in case of hospitalization, are hospitalized for less than 24 hours.

Highway pilot system characteristics and crash variables

Based on the highway pilot system description, four crash characteristics for the crashes addressed by the system were defined: 1) crashes involving at least one passenger vehicle; 2) crashes on highways, national roads and departmental roads with divided carriageway and at least two unidirectional lanes; 3) exclusion of crashes at an intersection; and 4) exclusion of crashes with heavy rain, snow, and hail. Additionally, the crash variables corresponding to the four crash characteristics of the BAAC database are illustrated in table 12.

Table 12 - Crash variables identified for a highway pilot system relative to crash variables in the BAAC database

Crash characteristic	Crash variables
Crashes involving at least one passenger vehicle	Type of vehicle
Crashes on highways, national roads and departmental roads with divided carriageways and at least two unidirectional lanes	Type of road network + Characteristics of the carriageway (divided or not, number of lanes, etc.)
Exclude crashes at an intersection	Location
Exclusion of crashes involving heavy rain, snow, and hail	Atmospheric conditions

Target population estimation

The crash characteristics corresponding to the scenarios in which the highway pilot system could potentially avoid crashes, were used to query the 2015 crashes coded in the BAAC database in order to estimate the target population of injury crashes and fatal crashes in terms of number of crashes, fatalities, injured and hospitalized road users, injured and not hospitalized road users. The results of the target population estimates are illustrated in table 13. Moreover, the estimates of the target population were compared relative to the annual frequencies of all injury crashes in 2015. The results of these comparisons (displayed in table 13) indicate that the highway pilot system could potentially address 4,6% of all injury accidents, and 3,8% of the road fatalities, 3,3% of the road users injured and hospitalized and 6,3% of the road users injured and not hospitalized.

Table 13 - Target population and target population relative to all crashes in 2015

	Number of crashes	Number of fatalities	Injured and hospitalized	Injured and not hospitalized
Injury crashes (target population)	2589	131	887	2779
Fatal crashes (target population)	117	131	73	64
Injury crashes (all crashes)	56603	3461	26595	44207
Target population relative to all crashes	4,6%	3,8%	3,3%	6,3%

3.4.3 Safety Requirements

The safety requirements related to the highway pilot system were identified by conducting an STPA analysis comprising four parts:

1. Establishing the system engineering foundation by defining the accidents, hazards and constraints of the system and by building the control structure of the highway pilot system.
2. Identification of the unsafe control actions of the system by performing STPA step 1. Additionally, a graphical timeline was created to illustrate the distribution of the unsafe control actions across the operation of the system (including driving modes and transitions).
3. Translating the identified unsafe control actions into safety requirements.
4. Generating scenarios leading to unsafe control actions and using the scenarios to define additional refined safety requirements via STPA step 2. In this analysis, the whole set of identified unsafe control actions were grouped into six categories in order to decrease the number of inputs for the elaboration of scenarios and thus reducing the processing time of the analysis.

System engineering foundation for the analysis

Establishing the system engineering foundation for the analysis consists of defining the accidents, hazards and constraints of the system, and building the control structure.

System accidents

ACC-1: People die or get injured due to a vehicle collision.

ACC-2: Property damage due to a vehicle collision.

System hazards

H-1: The vehicle violates the safety distance to other road users or objects on the road.

H-2: The vehicle leaves the roadway.

System safety constraints

SC-1: The safety control structure must prevent the vehicle from violating the safety distance to other road users or objects on the road.

SC-2: The safety control structure must prevent the vehicle from leaving the roadway.

Control structure:

The control structure of the highway pilot system examined in the STPA analysis is displayed in figure 26; it shows the human driver controller, the automated controller, the HMI component, actuators and sensors, networks, the driving environment, and their interactions in terms of feedback (blue arrows) and control actions (black arrows). The human driver controller receives feedback on the driving environment via human perception and on automation via the HMI.

Further, the human driver provides three control actions through the vehicle actuators (steering wheel, acceleration and braking pedals): the control of the vehicle, the release of the vehicle control and ADS override. Additionally, the human driver controller also provides three control actions via commands in the HMI: the validation of ADS engagement, the validation of the takeover request and ADS disengagement.

In turn, the automated controller receives feedback from the driving environment, the vehicle and the human driver via vehicle sensors and external information via networks (e.g. work zone ahead, end of ADS compatible road, etc.). The automated controller provides six control actions via the vehicle actuators: the engagement and disengagement of the ADS, the control of the vehicle, the release of vehicle control, the following of traffic rules and social norms, and the execution of minimal risk maneuvers. Lastly, the automated controller also provides two control actions via the HMI: sending the ADS availability notification and takeover requests, which may influence the human driver controller to validate ADS engagement and to validate takeover requests.

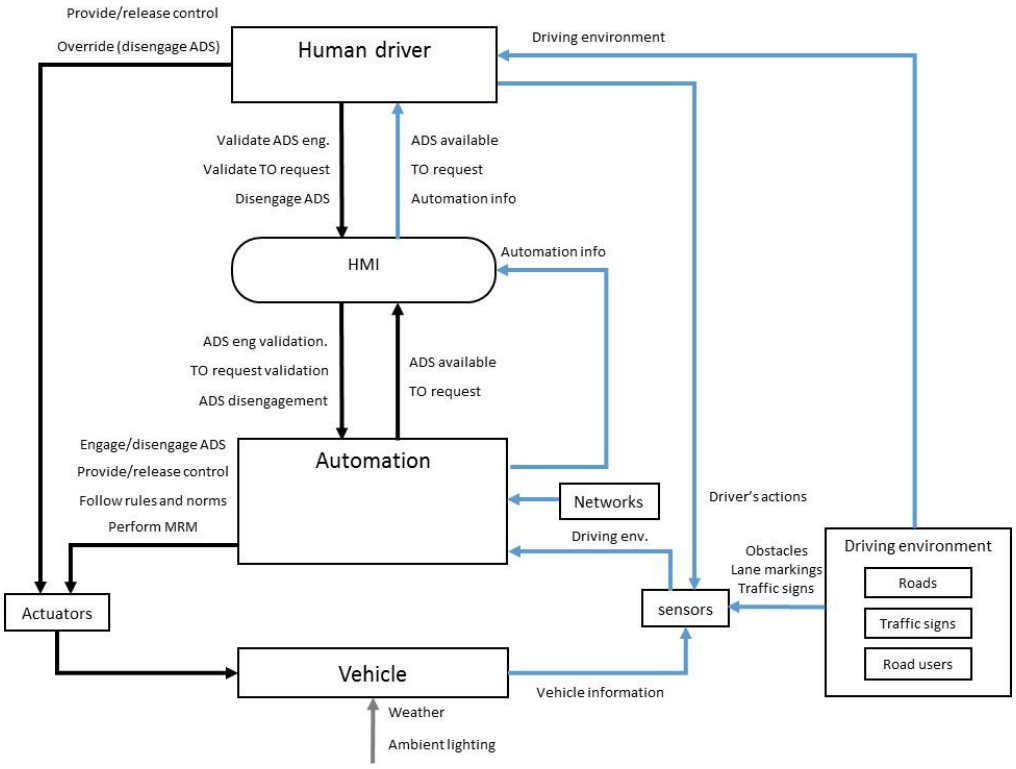


Figure 26 – Control structure of the highway pilot system

Unsafe control actions (STPA step 1)

In the first step of the STPA analysis, the control actions defined in the figure 26 (six for the human driver controller and eight for the automated controller) were examined according to the four types of unsafe control actions and documented using UCA tables, to identify contexts in which the control actions become unsafe control actions.

Table 14 illustrates some of the unsafe control actions identified for the human driver controller relative to the four types of unsafe control actions. For example, the control action in which the driver provides control of the vehicle can be unsafe when the driver provides inadequate control of the vehicle during manual driving due to the behavioral adaptation to automation (UCA-1), or when the driver does not provide control of the vehicle after the validation of a takeover request (UCA-23).

Furthermore, the control action in which the driver releases the control of the vehicle can be unsafe when the driver releases control of the vehicle too soon before the ADS is engaged. Finally, the control action in which the driver validates a takeover request sent by automation

can be unsafe if the driver does not validate the takeover request when automation sends the request (UCA-18) and if the driver validates the takeover request and puts the vehicle in an unsafe situation (UCA-19).

Table 14 - Example of UCA table containing unsafe control actions identified for the human driver controller

Hazards: Violating safety distance and leaving roadway				
Control action (CA)	**Not providing the CA causes hazard**	**Providing the CA causes hazard**	**Providing the CA too early/too late/wrong order causes hazard**	**Stopping the CA too soon/ applying the CA for too long causes hazard**
Provide control of the vehicle		**UCA-1:** Driver provides inadequate control of the vehicle during manual driving		
	UCA-23: Driver does not provide control of the vehicle after the validation of a takeover request			
Release control of the vehicle				**UCA-11:** Driver releases control of the vehicle too soon before the ADS is engaged
Validate takeover request	**UCA-18:** Driver does not validate takeover request when automation sends the request	**UCA-19:** Driver validates takeover request and puts the vehicle in an unsafe situation		

Table 15 displays some of the unsafe control actions identified for the automated controller. For instance, the control action in which the automated controller provides control of the vehicle can be unsafe when automation provides control of the vehicle during manual driving (UCA-2), when automation does not provide control of the vehicle after ADS engagement (UCA-8), and when automation provides inadequate control of the vehicle when ADS is engaged (UCA-9). Additionally, the control action in which the automated controller sends a takeover request is unsafe if automation does not send a takeover request when ADS conditions are no longer met i.e. ADS end mode type 2 (UCA-16).

Table 15 - Example of UCA table containing unsafe control actions identified for the automated controller

Hazards: Violating safety distance and leaving roadway				
Control action (CA)	**Not providing the CA causes hazard**	**Providing the CA causes hazard**	**Providing the CA too early/too late/wrong order causes hazard**	**Stopping the CA too soon/ applying the CA for too long causes hazard**
Provide control of the vehicle		**UCA-2:** Automation provides control of the vehicle during manual driving		
	UCA-8: Automation does not provide control of the vehicle after ADS engagement	**UCA-9:** Automation provides inadequate control of the vehicle when ADS is engaged		
Send takeover request	**UCA-16:** Automation does not send takeover request when the ADS conditions are no longer met (ADS end mode type 2)			

Overall, the first step of STPA identified 11 unsafe control actions for the human driver controller and 21 unsafe control actions for the automated driving controller; these 32 unsafe control actions are illustrated in table 16.

Table 16 – Unsafe control actions identified for the highway pilot system

UCA	UCA description
UCA-1	Driver provides inadequate control of the vehicle during manual driving
UCA-2	Automation provides control of the vehicle during manual driving
UCA-3	Automation sends ADS availability notification when ADS is not available
UCA-4	Driver provides ADS validation when it is inappropriate to engage the ADS
UCA-5	Automation does not engage ADS when the driver validates ADS engagement
UCA-6	Automation engages ADS when ADS engagement conditions are not met
UCA-7	Automation engages ADS when the driver does not validate ADS engagement
UCA-8	Automation does not provide control of the vehicle after ADS engagement
UCA-9	Automation provides inadequate control of the vehicle when ADS is engaged
UCA-10	Driver does not release the control of the vehicle after ADS engagement
UCA-11	Driver releases the control of the vehicle too soon before ADS engagement
UCA-12	Driver disengages ADS and puts the vehicle in an unsafe situation
UCA-13	Automation provides control of the vehicle after ADS conditions are no longer met
UCA-14	Automation follows traffic rules and/or social norms in an inadequate fashion
UCA-15	Automation follows traffic rules and/or social norms and puts the vehicle in an unsafe situation
UCA-16	Automation does not send takeover request when ADS conditions are no longer met (ADS end mode type 2)
UCA-17	Automation does not send takeover request when ADS compatible road comes to an end (ADS end mode type 1)
UCA-18	Driver does not validate takeover request when automation sends the takeover request
UCA-19	Driver validates takeover request and puts the vehicle in an unsafe situation
UCA-20	Automation does not disengage ADS when the driver validates a takeover request
UCA-21	Automation disengages ADS when the driver has not validated a takeover request
UCA-22	Automation does not release control of the vehicle when the driver validates a takeover request
UCA-23	Driver does not provide control of the vehicle after the validation of a takeover request
UCA-24	Driver provides inadequate control of the vehicle after the validation of a takeover request
UCA-25	Automation does not provide minimal risk maneuver when the driver does not respond to the takeover request (end mode type 1 and end mode type 2)
UCA-26	Automation does not provide minimal risk maneuver when automation can no longer assure the safe operation of the vehicle
UCA-27	Automation provides minimal risk maneuver and puts the vehicle in an unsafe situation
UCA-28	Driver provides inadequate control of the vehicle after a minimal risk maneuver
UCA-29	Automation does not disengage ADS when the driver provides ADS disengagement
UCA-30	Automation disengages ADS when the driver has not provided ADS disengagement
UCA-31	Automation does not release control of the vehicle when the driver provides ADS disengagement
UCA-32	Driver provides inadequate control of the vehicle after ADS disengagement

Timeline

A timeline (figure 27) containing the control actions (displayed on the top) and the unsafe control actions (displayed at the bottom) was created in order to assist the understanding of the distribution of unsafe control actions relative to the five phases of the highway pilot

system's operation. Moreover, the elements in blue are related to automation, and the elements in green are related to the human driver.

- **Phase A:** It involves the manual driving stage before the engagement of the ADS;
- **Phase B:** It comprises the transition from manual driving to automated driving which is triggered after the driver's validation of ADS engagement;
- **Phase C:** It encompasses the automated driving stage;
- **Phase D:** It corresponds to the stage of the transition from automated driving to manual driving initiated by automation which can happen in three ways (i.e. three end modes). Accordingly, phase D is divided in three sub-phases:
    - **Phase D.1:** Automation sends a takeover request (ADS end modes 1 and 2) and the driver validates the request;
    - **Phase D.2:** The driver does not validate the takeover request or automation detects a performance –relevant failure (ADS end mode type 3), leading automation to perform a minimal risk maneuver; and
    - **Phase D.3:** The driver disengages ADS by overriding the ADS through actions on the steering wheel or pedals, or by providing the ADS disengagement command;
- **Phase E:** It comprises the stage in which the human driver performs manual driving after a transition from automated driving mode. As in phase D, it also happens in three ways, and therefore phase E is divided in three sub-phases:
    - **Phase E.1:** The driver performs manual driving after the validation of a takeover request;
    - **Phase E.2:** The driver performs manual driving after a minimal risk maneuver executed by automation; and
    - **Phase E.3:** The driver performs manual driving after an ADS disengagement initiated by the driver.

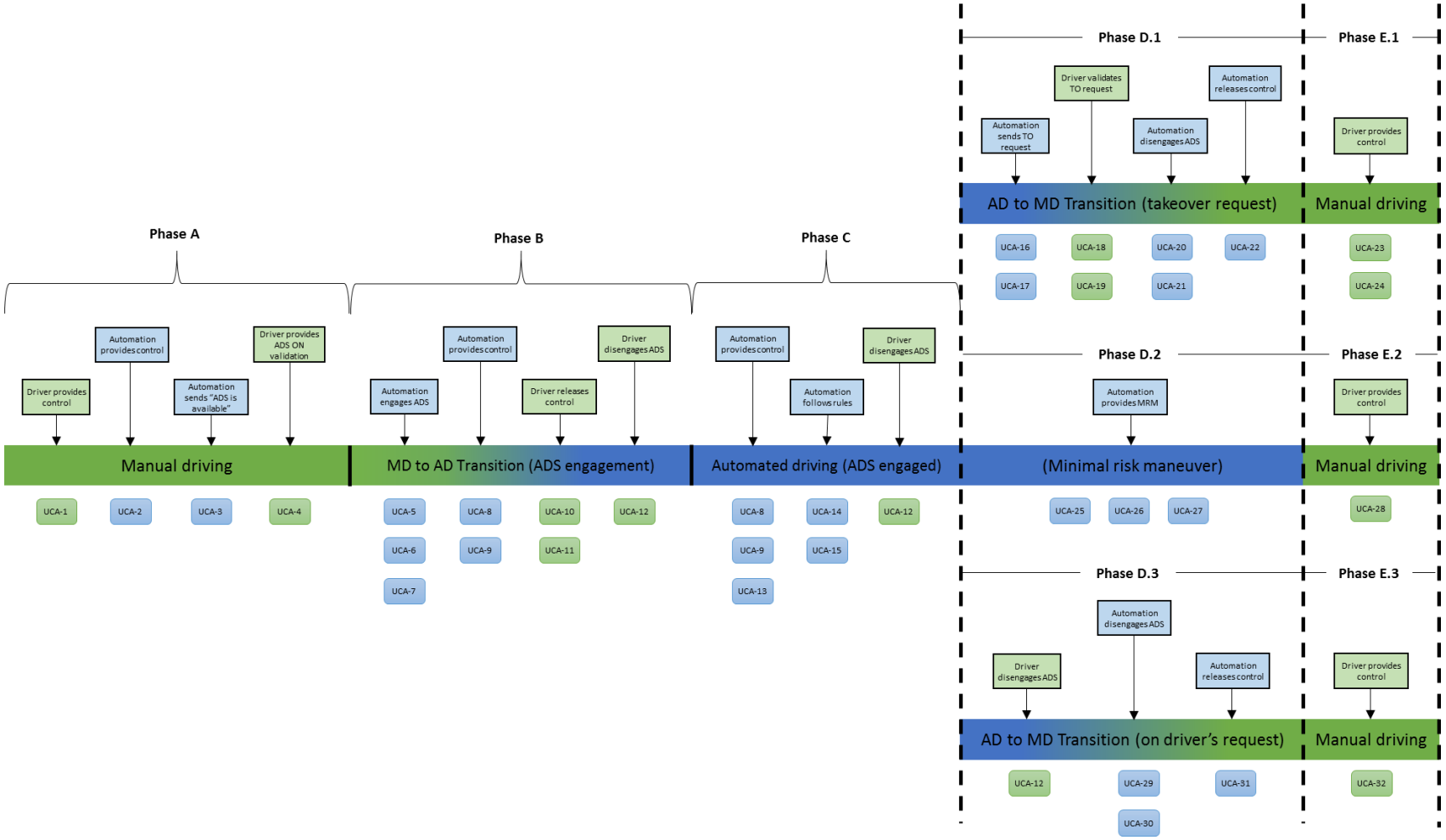


Figure 27 - Timeline relative to the control actions and unsafe control actions of the highway pilot system

The timeline displays the distribution of the control actions (on the top) and unsafe control actions (on the bottom) of the timeline across the five phases of the highway pilot system's operation (A-E). The green color is related to the human driver and manual driving and the blue to automation and automated driving.

Safety requirements

The unsafe control actions identified in the first step of the STPA analysis are used to define safety requirement on the component behavior. For example, the unsafe control actions illustrated in table 16 can be translated into the following safety requirements:

- **UCA-1:** The driver provides inadequate control of the vehicle during manual driving.
- **SR-1:** The driver must provide adequate control of the vehicle during manual driving.

- **UCA-11:** The driver releases control of the vehicle too soon before the ADS is engaged.
- **SR-11:** The driver must not release control of the vehicle too soon before the ADS is engaged.

- **UCA-18:** The driver does not validate the takeover request when automation sends the request.
- **SR-18:** The driver should[8] validate the takeover request when automation sends a takeover request.

- **UCA-19:** The driver validates takeover request and puts the vehicle in an unsafe situation.
- **SR-19:** The driver must not put the vehicle in an unsafe situation after the validation of a takeover request.

- **UCA-23:** The driver does not provide control of the vehicle after the validation of a takeover request.
- **SR-23:** The driver must provide control of the vehicle after the validation of a takeover request.

[8] The word should is used in SR-18 because there is a fallback performance strategy in which automation executes a minimal risk maneuver when the driver does not validate the takeover request. However, it is safer when the driver regains situational awareness, validates the takeover request and provides adequate control of the vehicle.

The whole set of safety requirements defined using the unsafe control actions from table 16, are displayed in table 17.

Table 17 - Safety requirements defined for the highway pilot system using the unsafe control actions

SR	Safety requirement description
SR-1	Driver must provide adequate control of the vehicle during manual driving
SR-2	Automation must not provide control of the vehicle during manual driving
SR-3	Automation must not send ADS is availability notification when ADS is not available
SR-4	Driver must not provide ADS validation when it is inappropriate to engage the ADS
SR-5	Automation must engage ADS when the driver validates ADS engagement
SR-6	Automation must not engage ADS when ADS engagement conditions are not met
SR-7	Automation must not engage ADS when the driver does not validate ADS engagement
SR-8	Automation must provide control of the vehicle after ADS engagement
SR-9	Automation must provide adequate control of the vehicle when ADS is engaged
SR-10	Driver must release control of the vehicle after ADS engagement
SR-11	Driver must not release control of the vehicle too soon before ADS engagement
SR-12	Driver must not put the vehicle in an unsafe situation when s/he disengages ADS
SR-13	Automation must not provide control of the vehicle after ADS conditions are no longer met
SR-14	Automation must follows traffic rules and/or social norms in an adequate fashion
SR-15	Automation must not put the vehicle in an unsafe situation when automation follows traffic rules and/or social norms
SR-16	Automation must send takeover request when ADS conditions are no longer met (ADS end mode type 2)
SR-17	Automation must send takeover request when ADS compatible road comes to an end (ADS end mode type 1)
SR-18	Driver should validate takeover request when automation sends the takeover request
SR-19	Driver must not put the vehicle in an unsafe situation when s/he validates takeover request
SR-20	Automation must disengage ADS when the driver validates a takeover request
SR-21	Automation must not disengage ADS when the driver has not validated a takeover request
SR-22	Automation must release control of the vehicle when the driver validates a takeover request
SR-23	Driver must provide control of the vehicle after the validation of a takeover request
SR-24	Driver must provide adequate control of the vehicle after the validation of a takeover request
SR-25	Automation must provide minimal risk maneuver when the driver does not respond to the takeover request (end mode type 1 and end mode type 2)
SR-26	Automation must provide minimal risk maneuver when automation can no longer assure the safe operation of the vehicle
SR-27	Automation must not put the vehicle in an unsafe situation when automation provides minimal risk maneuver and
SR-28	Driver must provide adequate control of the vehicle after a minimal risk maneuver
SR-29	Automation must disengage ADS when the driver provides ADS disengagement
SR-30	Automation must not disengage ADS when the driver has not provided ADS disengagement
SR-31	Automation must release control of the vehicle when the driver provides ADS disengagement
SR-32	Driver must provide adequate control of the vehicle after ADS disengagement

Scenarios and refined safety requirements (STPA step 2)

In the second step of the STPA analysis, the identified unsafe control actions are analyzed to generate scenarios (i.e. identifying potential causes) leading to unsafe control actions and to create refined safety requirements. Although a first attempt was made to generate scenarios for the 32 identified unsafe control actions, the generated scenarios showed that multiple unsafe control actions had the same potential causes. For example, inaccurate measurements provided by vehicle sensors on the ADS conditions can cause automation to send an ADS availability notification when the ADS is not available and not to send a takeover request when the ADS conditions are no longer met. Consequently, the 32 unsafe control actions were classified into six categories in order to optimize the analysis processing time in the elaboration of scenarios.

Classification of the unsafe control actions

The timeline displayed in figure 27 was analyzed to identify the unsafe control actions with similar nature for the two controllers. For example, there are several unsafe control actions provided by automation to send notifications or requests to the driver, and multiple unsafe control actions related to automation's execution of the vehicle control. Conversely, there are various unsafe control actions associated to the driver's response to feedback provided by automation, and a few control actions related to the driver's execution of the vehicle control.

The result of this analysis enabled to establish six categories of unsafe control actions which are illustrated in figure 28. The right side of the figure shows three categories in blue which are associated to the automated controller, and the left side shows three categories in green which are related to the human driver controller. At the bottom of every category, the unsafe control actions contained in the category are illustrated in grey boxes. Moreover, in the center of the figure, the control structure of the highway pilot system is displayed with numbers indicating the control loops related to each category.

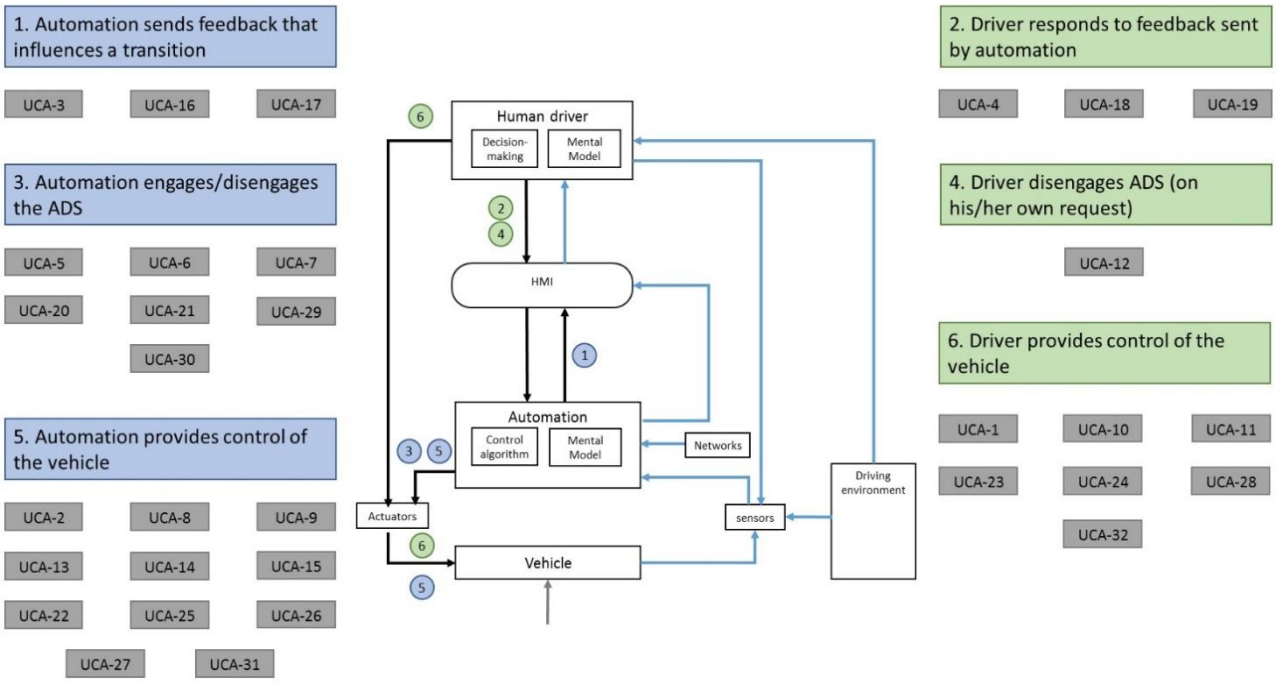


Figure 28 - Classification of the unsafe control actions

The categories of unsafe control actions related to the automated controller are displayed in blue and the ones related to the human driver are displayed in green. The unsafe control actions contained in each category are illustrated in grey boxes. The numbers corresponding to the categories are distributed on the control structure in to indicate the part of the control loop associated to each category.

1. Automation sends feedback to influence the driver to initiate a transition

The first category encompasses unsafe control actions related to control actions in which automation sends feedback on ADS availability notification and takeover requests in order to influence the driver to initiate a driving mode transition. The ADS availability notification may influence the driver to initiate a transition from manual driving to automated driving. Additionally, takeover requests may influence the driver to validate the takeover request, and to initiate a transition from automated driving to manual driving.

2. Driver responds to feedback sent by automation to influence a transition

The second category includes three unsafe control actions associated to the control actions involving the driver's response to the feedback that automation sends to influence a transition. As aforementioned, the driver's response to ADS availability notification via the ADS engagement validation command triggers a transition from manual driving to automated

driving. Also, the driver's response to takeover requests via the takeover validation command, triggers a transition from automated driving mode to manual driving mode.

3. Automation engages/disengages the automated driving system

The third category covers the unsafe control actions related to automation's control actions regarding ADS engagement and ADS disengagement. After the driver's ADS engagement validation, automation still has to verify ADS conditions before engaging the ADS. Further, automation disengages the ADS in three contexts: the driver's validation of a takeover request, driver's override of the ADS and driver's ADS disengagement command.

4. Driver disengages the automated driving system on driver's request

The fourth category contains one unsafe control action in which the driver initiates ADS disengagement on his/her own request, that is, the driver provides ADS disengagement without a takeover request. The driver can initiate an ADS disengagement by overriding the system when drivers provide acceleration, braking or steering, or by providing the ADS disengagement command.

5. Automation provides control of the vehicle

The fifth category comprises the unsafe control actions associated to control actions provided by automation which require to control the vehicle via the vehicle's actuators: a) provide control of the vehicle; b) follow traffic rules and social norms; c) release control of the vehicle; and d) perform the minimal risk maneuver.

6. Driver provides control of the vehicle

The last category involves the unsafe control actions related to the control actions provided by the driver to control the vehicle and to release the control of the vehicle.

Scenarios and refined safety requirements

The six categories of unsafe control actions were examined using Leveson's control flaw classification (figure 29) to elaborate 53 scenarios. In turn, the elaborated scenarios were used to define 80 refined safety requirements. Moreover, the control flaw classification in figure 29 was aggregated into four high-level classes related to feedback and inputs, models, decision-

making and action execution, to organize the scenarios and refined safety requirements in tables according to the four high-level classes.

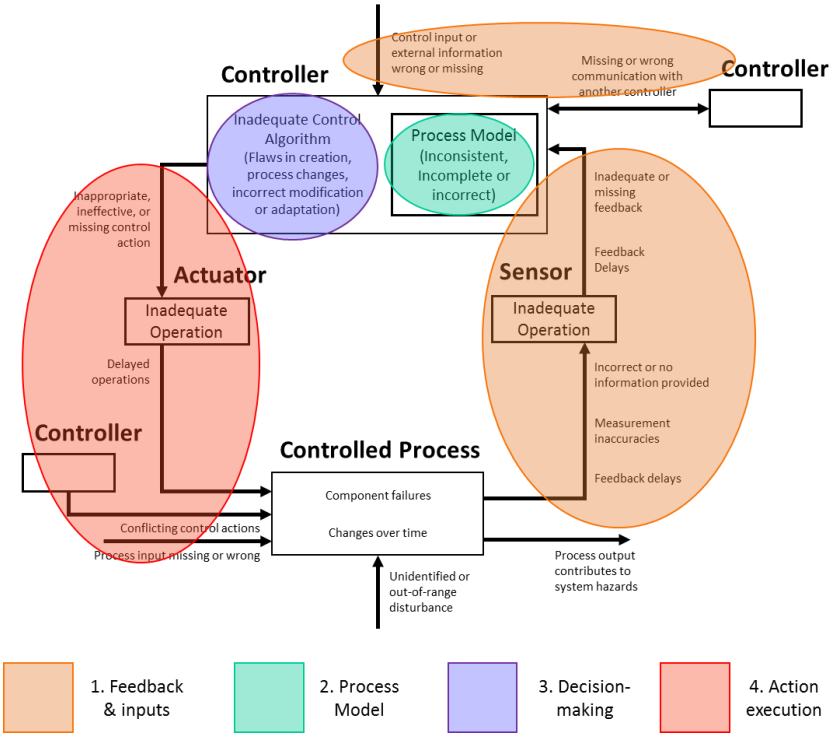


Figure 29 – High-level control flaws classes related to the control structure

The control flaws proposed by Leveson were aggregated into four high-level classes according to color-codes. The orange color indicates control flaws related to feedback and inputs, the blue color to those related to process models, purple to those related to decision-making and red to those related to action execution.

To illustrate the approach, the scenarios and safety requirements for two categories of unsafe control actions (categories one and two) are presented below. The complete results for all the six categories can be viewed in Appendix A.

Scenarios and refined safety requirements in category 1—automation sends feedback to influence the driver to initiate a transition

The unsafe control actions encompassed in category 1, were examined according to the control flaw classification (figure 29) to generate nine scenarios and 11 refined safety requirements. The results of the analysis were organized according to the four high-level classes.

Feedback and inputs:

Automation receives feedback and inputs (e.g. on ADS availability conditions, ADS conditions, driving environment, automation's performance and the driver's actions) from vehicle sensors and external information such as networks, to help automation determine when ADS is available and when a takeover request is necessary.

Scenario 1: Automation sends ADS availability notification when ADS is not available or does not send ADS takeover request when needed, due to measurement inaccuracies, feedback delays or no information measured by vehicle sensors on necessary feedback to determine ADS availability and the need for a takeover request.

- **RSR-1:** Vehicle sensors must take accurate on-time measures on the necessary feedback to determine ADS availability and the need for a takeover request.
- **RSR-2:** Automation must detect when vehicle sensors are providing inaccurate measures with delays of TBD, on the necessary feedback to determine ADS availability and the need for a takeover request.

Scenario 2: Automation sends ADS availability notification when ADS is not available or does not send ADS takeover request when needed, due to the inadequate operation of vehicle sensors which cause incorrect feedback regarding ADS availability or the need for a takeover request.

- **RSR-3:** Vehicle sensors that measure the necessary feedback to determine ADS availability and the need for a takeover request, must have an adequate operation.
- **RSR-4:** Automation must detect when the vehicle sensors that measure necessary feedback to determine ADS availability and the need for a takeover request**,** have an inadequate operation.

Scenario 3: Automation sends ADS availability notification when ADS is not available or does not send ADS takeover request when needed, due to missing or inadequate feedback provided by vehicle sensors regarding the necessary feedback to determine ADS availability and the need for a takeover request**.**

- **RSR-5:** Vehicle sensors must provide adequate feedback on the necessary information to determine ADS availability and the need for a takeover request**.**

Scenario 4: Automation sends ADS availability notification when ADS is not available or does not send ADS takeover request when needed, due to missing or inadequate feedback provided by external information (e.g. networks).

- **RSR-6:** External information (e.g. networks) must provide adequate feedback on the feedback necessary to determine ADS availability and the need for a takeover request.

Models:

Even if automation receives adequate feedback from vehicle sensors and external information, automation still has to build appropriate models regarding ADS availability and the need for takeover requests, which the control algorithm uses to generate the ADS availability notification and takeover requests. Models include the conditions that ADS designers have selected for the ADS availability (e.g. ADS compatible road, speed range, surrounding traffic characteristics, etc.) and the need for takeover requests (end of ADS compatible road, problem with the perception system, etc.). Moreover, these conditions are translated into software requirements which are included into automation's algorithm, and feedback inputs are defined in order to enable automation evaluate conditions on ADS availability and the need for takeover requests. For example, an ADS condition defined by designers as heavy traffic, has to be translated into a requirement that automation can apply such as "TBD% of the time of the last TBD seconds, another vehicle has been present on each adjacent lane, ahead or abeam the EGO vehicle, with a time gap to the EGO vehicle of less than TBD seconds". Also, the feedback inputs that allow automation to evaluate the presence of other vehicles are defined e.g. radar and camera signals for the detection of other vehicles.

As a result, two scenarios were defined and two refined safety requirements were established for automation's models:

Scenario 5: Automation sends ADS availability notification when ADS is not available or does not send ADS takeover request when needed, due to an inadequate model of ADS conditions.

- **RSR-7:** Automation must have an adequate model of ADS availability conditions and an adequate model of ADS conditions to continue on automated driving.

Scenario 6: Automation sends ADS availability notification when ADS is not available or does not send ADS takeover request when needed, because automation is not aware that ADS is not available or that a takeover request is needed due to an inadequate model of the state of ADS conditions.

- **RSR-8:** The software requirements and feedback inputs included in automation's model must enable automation to adequately assess the state of ADS conditions.

Decision-making (control algorithm):

Automation's control algorithm should generate control actions based on automation's models of the controlled process. However, flaws in software requirements and "software errors", can lead the control algorithm to generate ADS availability notifications, when automation's model is aware that ADS is not available, and to not generate takeover requests when automation's model is aware that a takeover request is necessary.

Scenario 7: Automation sends ADS availability notification when ADS is not available or does not send ADS takeover request when needed, due to flaws in software requirements and software errors in automation's control algorithm.

- **RSR-9:** Automation's control algorithm must not generate ADS availability notification when the model indicates ADS is not available, and must generate takeover requests when the ADS conditions are no longer met.

Action execution:

Once the control algorithm has generated the ADS availability notification or a takeover request, a signal has to be sent to the HMI to display these feedback to the driver. The issues in the execution of these control actions may include not sending the signal (missing control action) or sending the signal with delays.

Scenario 8: Automation sends ADS availability notification when ADS is not available or does not send ADS takeover request when needed, because the signal is not sent to the HMI due to problems with communication.

- **RSR-10:** Automation must ensure that the actions generated by the control algorithm to send the ADS availability notification and the takeover requests to the HMI, are executed.

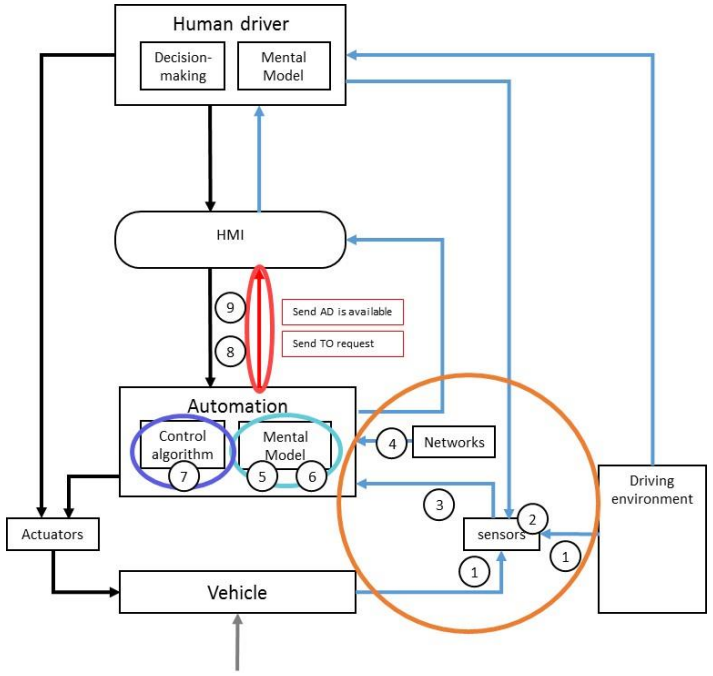
Scenario 9: Automation sends ADS availability notification when ADS is not available or does not send ADS takeover request when needed, because the signal is sent with delays to the HMI, or due to problems with communication.

- **RSR-11:** Automation must ensure that the actions generated by the control algorithm to send ADS availability notification and takeover requests to the HMI are sent with a maximal delay of TBD.

The results of the STPA analysis for category one are illustrated in table 18. The first part of the table shows the unsafe control actions and safety requirements contained in category one, which were identified in the step 1 of the STPA analysis. The second part of the table includes the part of the control loop that was considered in the analysis, the high-level classes and control flaws associated to the nine scenarios and the 11 refined safety requirements. Moreover, the table displays the control structure of the highway pilot system and the control loops examined in the analysis with a color code that indicates the high-level classes.

Table 18 - Synthesis of STPA results for category 1

The first part of the table displays the three unsafe control actions and the three corresponding safety requirements contained in the category 1. The second part of the table illustrates the scenarios elaborated for the category 1 structured according to the four high-level classes and their color codes. The control structure on the left circles the part of the control loop concerned by the high-level classes using the color codes and the specific control flaws using numbers.

<table>
<tr><th colspan="4">CATEGORY 1: Automation sends feedback to influence a transition</th></tr>
<tr><th colspan="2">STPA step 1</th><th colspan="2">UCAs translated into safety requirements</th></tr>
<tr><th colspan="2">Unsafe control actions</th><th colspan="2">Safety requirements</th></tr>
<tr><td colspan="2">UCA-3: Automation sends ADS availability notification when ADS is not available</td><td colspan="2">SR-3: Automation must not send ADS availability notification when ADS is not available</td></tr>
<tr><td colspan="2">UCA-16: Automation does not send takeover request when the ADS conditions are no longer met (end mode type 2)</td><td colspan="2">SR-16: Automation must send takeover request when the ADS conditions are no longer met (end mode type 2)</td></tr>
<tr><td colspan="2">UCA-17: Automation does not send takeover request when the ADS compatible road comes to an end e.g. highway exit (end mode type 1)</td><td colspan="2">SR-17: Automation must send takeover request when the ADS compatible road comes to an end e.g. highway exit (end mode type 1)</td></tr>
<tr><th colspan="4">STPA step 2</th></tr>
<tr><th>Control structure</th><th>Class</th><th>Control flaw</th><th>Refined safety requirement</th></tr>
<tr><td rowspan="10"> 
</td><td rowspan="6">Feedback and inputs</td><td rowspan="2">Measurement inaccuracies, feedback delays or no information measured by vehicle sensors (1)</td><td>RSR-1: Vehicle sensors must take accurate on time measures on the necessary feedback to determine that ADS is available and that a takeover request is needed</td></tr>
<tr><td>RSR-2: Automation must detect when vehicle sensors are providing inaccurate measures with delays of TBD, on the necessary feedback to determine that ADS is available and that a takeover request is needed</td></tr>
<tr><td rowspan="2">Inadequate sensor operation (2)</td><td>RSR-3: Vehicle sensors that measure the necessary feedback to determine that ADS is available and that a takeover request is needed, must have an adequate operation</td></tr>
<tr><td>RSR-4: Automation must detect when the vehicle sensors that provide the necessary feedback to determine that ADS is available and that a takeover request is needed, have an inadequate operation</td></tr>
<tr><td>Missing or inadequate feedback on ADS conditions sent by vehicle sensors (3)</td><td>RSR-5: Vehicle sensors must provide adequate feedback on the necessary information to determine that ADS is available and that a takeover request is needed, have an inadequate operation</td></tr>
<tr><td>Missing or inadequate external information on ADS conditions (4)</td><td>RSR-6: External information (e.g. networks) must provide adequate information on the necessary feedback to determine that ADS is available and that a takeover request is needed, have an inadequate operation</td></tr>
<tr><td rowspan="2">Model</td><td>Inadequate model of the state of ADS conditions (5)</td><td>RSR-7: Automation must have an adequate model of ADS availability conditions and an adequate model of ADS conditions to continue on automated driving</td></tr>
<tr><td>Inadequate model of ADS conditions (6)</td><td>RSR-8: The software requirements and feedback inputs included in automation’s model must enable automation to adequately assess the state of ADS conditions</td></tr>
<tr><td>Decision-making</td><td>Inadequate control algorithm (7)</td><td>RSR-9: Automation’s control algorithm must not generate ADS availability notification when the model indicates ADS is not available, and must generate takeover requests when the ADS conditions are no longer met</td></tr>
<tr><td rowspan="2">Action execution</td><td>Missing control action (8)</td><td>RSR-10: Automation must ensure that the actions generated by the control algorithm to send the ADS availability notification and the takeover requests to the HMI, are executed</td></tr>
<tr><td>Delayed operation (9)</td><td>RSR-11: Automation must ensure that the actions generated by the control algorithm to send the ADS availability notification and the takeover requests to the HMI are sent with a maximal delay of TBD</td></tr>
</table>

Scenarios and refined requirements in category 2—the driver responds to feedback sent by automation to influence a transition

The unsafe control actions included in category 2 were examined according to the control flaw classification to generate nine scenarios and 16 refined safety requirements. As in category one, the results of the analysis were organized according to the four high-level classes.

Feedback and inputs:

The driver perceives feedback sent by automation to initiate transitions via the HMI (i.e. ADS is available notification and takeover requests) and feedback on the driving environment to determine when it is appropriate to engage the ADS, and to regain situation awareness before the validation of the takeover request.

Scenario 10: The driver does not validate the takeover request because the feedback on the HMI is missing due to a problem in communication or inadequate operation of HMI components, and therefore, the takeover request is never displayed on the HMI.

- **RSR-12:** There must be an adequate communication between automation and the HMI, and an adequate HMI operation that enables displaying feedback provided by automation on takeover requests.

Scenario 11: The driver does not validate takeover request because s/he does not perceive the feedback on the HMI due to inadequate feedback (e.g. inconsistent feedback, difficult to perceive, difficult to understand, etc.) displayed on the HMI.

- **RSR-13:** The HMI must provide adequate feedback to the driver on ADS availability notification and takeover requests.
- **RSR-14:** The mental model of the driver must include the procedures and knowledge necessary to understand the feedback provided by the HMI.
- **RSR-15:** The driver must value being receptive to the feedback provided by the HMI.

Scenario 12: The driver validates ADS engagement in inappropriate situations or puts the vehicle in an unsafe situation after validating a takeover request, because s/he does not perceive feedback on the driving environment.

- **RSR-16:** The driver must be able to perceive and detect the aspects that make it inappropriate to engage the ADS.
- **RSR-17:** The takeover procedures must enable the driver to perceive the traffic environment before the validation of the takeover request.

Models:

After the perception of feedback, drivers must update their mental models on whether or not it is appropriate to engage the ADS, on the takeover request and on the driving environment. The mental models of the drivers must include the knowledge and procedures necessary to determine when it is inappropriate to engage the ADS, and to be able to safely respond to a takeover request.

Scenario 13: The driver does not validate takeover request or validates the takeover request and puts the vehicle in an unsafe situation because s/he does not have an adequate model of the takeover procedure.

- **RSR-18:** The mental model of the driver must include knowledge on the takeover procedures.
- **RSR-19:** The procedures to validate a takeover request must be intuitive and easy to be performed by the driver.
- **RSR-20:** The HMI must provide adequate feedback to the driver on the steps to validate a takeover request.

Scenario 14: The driver validates ADS engagement when it is inappropriate or validates the takeover request and puts the vehicle in an unsafe situation because s/he does not have an adequate model of the driving environment.

- **RSR-21:** The mental model of the driver must include the situations when it is inappropriate to engage ADS
- **RSR-22:** The driver must have an adequate model of the traffic environment before the validation of the ADS engagement and takeover requests.

Decision-making:

Even if the driver has adequate models of the situations in which it is inappropriate to engage the ADS, on how to respond to a takeover request and on the driving environment, the driver may decide to engage the ADS when in it is inappropriate, or to systematically wait for the minimal risk maneuver instead of regaining situation awareness and validating the takeover request.

Scenario 15: The driver validates ADS engagement when it is inappropriate or does validate takeover requests because s/he decides that automation can handle inappropriate situations or that s/he can always rely on automation's minimal risk maneuver.

- **RSR-23:** The mental model of the driver must include safety values that encourage an adequate decision-making process regarding ADS engagement and takeover request validations.

Action execution:

The driver may provide unintended actions regarding the ADS engagement validation and takeover request validation, e.g. unintendedly pushing the button(s) that provides validation commands. Also, the driver may not be familiar with the procedures to provide validations (the sequences, the order, the place of the commands on the table board, etc.). Finally, the driver may provide commands for ADS engagement validation and takeover request validation, but the signals may not reach automation due to inadequate command operation or problems with communications.

Scenario 16: The driver validates ADS engagement when it is inappropriate or validates a takeover request and puts the vehicle in an unsafe situation because the driver provides an unintended validation of the commands on the HMI.

- **RSR-24:** The procedures and commands to validate ADS engagement and takeover requests must limit unintended validations.

Scenario 17: The driver does not validate a takeover request because s/he does not know the procedure to validate it (sequences, order, when, command location, etc.) and thus is unable to provide the takeover request validation.

- **RSR-25:** The mental model of the driver must include the location of the validation commands, the sequences, order, etc.
- **RSR-26:** The design of the validation commands and the HMI display information with takeover request must assist the driver to safely validate takeover requests.

Scenario 18: The driver provides ADS engagement validation command and takeover validation command, but the control action is not provided to automation because there is an inadequate command operation or problems with communication.

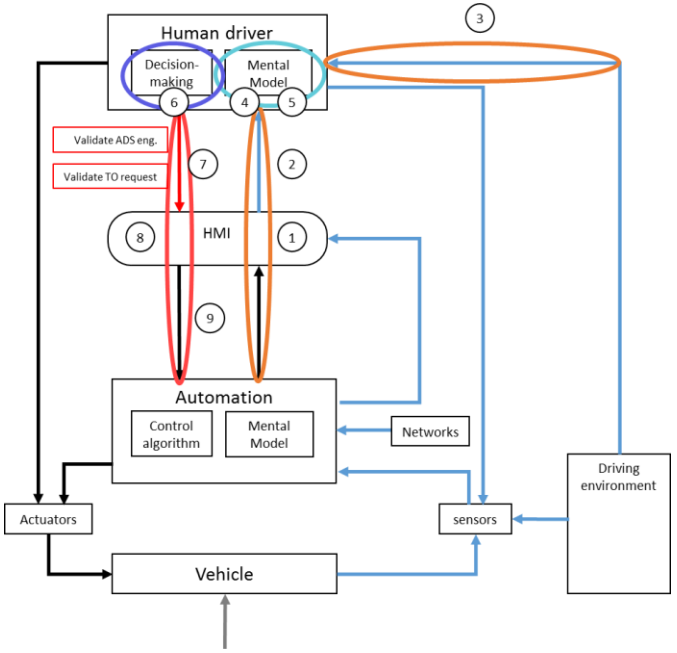
- **RSR-27:** The HMI commands must have an adequate operation and there must be an adequate communication between the HMI and automation, which ensures the actions provided by the driver reach automation.

As for category one, the results of category two, are displayed in table 19. The first part of the table shows the unsafe control actions and safety requirements contained in category two, which were identified in the step 1 of the STPA analysis. The second part of the table includes the control loops which were considered in the analysis, the classes, control flaws associated to the nine scenarios and the 16 refined safety requirements. Additionally, the table displays the control structure of the highway pilot system and the control loops examined in the analysis with a color code that indicates the control flaw classes.

The 32 safety requirements established using the unsafe control actions and the 80 refined safety requirements defined through the STPA step 2, provide the inputs for the next section in which safety requirements are used to derive questions that assist the evaluation of the assumptions related to direct safety mechanisms (1-2).

Table 19 - Synthesis of STPA results for category 2

The first part of the table displays the three unsafe control actions and the three corresponding safety requirements contained in the category 2. The second part of the table illustrates the scenarios elaborated for the category 2 structured according to the four high-level classes and their color codes. The control structure on the left circles the part of the control loop concerned by the high-level classes using the color codes and the specific control flaws using numbers.

<table>
<tr><th colspan="4">CAREGORY 2: Driver responds to feedback sent by automation</th></tr>
<tr><th colspan="2">STPA step 1</th><th colspan="2">UCAs translated into safety requirements</th></tr>
<tr><th colspan="2">Unsafe control actions</th><th colspan="2">Safety requirements</th></tr>
<tr><td colspan="2">UCA-4: Driver provides ADS validation when it is inappropriate to engage ADS</td><td colspan="2">SR-4: Driver must not provide ADS validation when it is inappropriate to engage ADS</td></tr>
<tr><td colspan="2">UCA-18: Driver does not validate the takeover request when automation sends the takeover request</td><td colspan="2">SR-18: Driver must validate the takeover request when automation sends the takeover request</td></tr>
<tr><td colspan="2">UCA-19: Driver validates takeover request and puts the vehicle in an unsafe situation</td><td colspan="2">SR-19: Driver must not put the vehicle in an unsafe situation after the validation of the takeover request</td></tr>
<tr><th colspan="4">STPA step 2</th></tr>
<tr><th>Control structure</th><th>Class</th><th>Control flaw</th><th>Refined safety requirement</th></tr>
<tr><td rowspan="10"> 
</td><td rowspan="6">Feedback and inputs</td><td>Inadequate or missing feedback provided by the HMI (1)</td><td>RSR-12: There must be an adequate communication between automation and the HMI, and an adequate HMI operation that enables to display the feedback provided by automation on ADS availability notification and takeover requests.</td></tr>
<tr><td rowspan="3">Inadequate human perception on the HMI (2)</td><td>RSR-13: The HMI must provide adequate feedback to the driver on ADS availability notification and takeover requests.</td></tr>
<tr><td>RSR-14: The mental model of the driver must include the procedures and knowledge necessary to understand the feedback provided by the HMI.</td></tr>
<tr><td>RSR-15: The driver must value being receptive to the feedback provided by the HMI</td></tr>
<tr><td rowspan="2">Inadequate human perception on the traffic environment (3)</td><td>RSR-16: The driver must be able to perceive and detect the aspects that make it inappropriate to engage the ADS</td></tr>
<tr><td>RSR-17: The takeover procedures must enable the driver to perceive the traffic environment before the validation of the takeover request</td></tr>
<tr><td rowspan="4">Model</td><td rowspan="3">Inadequate model of takeover request (4)</td><td>RSR-18: The mental model of the driver must include knowledge on the takeover procedures</td></tr>
<tr><td>RSR-19: The procedures to validate a takeover request must be intuitive and easy to perform by the driver</td></tr>
<tr><td>RSR-20: The HMI must provide adequate feedback to the driver on the steps to validate a takeover request</td></tr>
<tr><td rowspan="2">Inadequate model of the driving environment (5)</td><td>RSR-21: The mental model of the driver must include the situations when it is inappropriate to engage ADS</td></tr>
<tr><td>RSR-22: The driver must have an adequate model of the traffic environment before the validation of the ADS engagement and takeover requests</td></tr>
</table>

STPA step 2			
Control structure	**Class**	**Control flaw**	**Refined safety requirement**
	Decision-making	Inadequate control algorithm (6)	**RSR-23:** The mental model of the driver must include safety values that encourage an adequate decision-making process regarding ADS engagement and takeover request validations
	Action execution	Inappropriate control action (7)	**RSR-24:** The procedures and commands to validate ADS engagement and takeover requests must limit unintended validations
		Missing control action (7) and (8)	**RSR-25:** The mental model of the driver must include the location of the validation commands, the sequences, order, etc.
			RSR-26: The design of the validation commands and the HMI display information with takeover request must assist the driver to safely validate takeover requests
		Inadequate actuator operation and communication (8) and (9)	**RSR-27:** The HMI commands must have an adequate operation and there must be an adequate communication between the HMI and automation, which ensures the actions provided by the driver reach automation

3.4.4 Questions to consider in the evaluation of direct mechanisms (1-2)

The first step to derive questions to assist the evaluation of direct mechanisms was to assign which of the 32 safety requirements and 80 refined safety requirements addressed the direct safety mechanism one (direct impacts that vehicle systems have on the driving task), the direct safety mechanism two (direct influence of infrastructure on the driving task), or both.

Table 20 illustrates the process followed to assign safety requirements to the direct safety mechanisms. For each category, the safety requirements based on unsafe control actions were and the refined safety requirements were analyzed and assigned to the direct safety mechanisms.

Table 20 - Example of the allocation of safety requirements from category 1 to safety mechanisms

The first part of the table displays the safety requirements contained in category 1 which were assigned to safety mechanisms 1 and 2. The second part of the table shows the refined safety requirement organized according to the four high-level classes, which were assigned to safety mechanisms 1 and 2.

Category 1: Automation sends feedback to influence a transition			
Safety requirements (based on UCAs)		**SM-1**	**SM-2**
SR-3: Automation must not send ADS availability notification when ADS is not available		X	X
SR-16: Automation must not send takeover requests when the ADS conditions are no longer met		X	X
SR-17: Automation must not send takeover requests when the ADS compatible road comes to an end		X	X
Class	**Refined safety requirements (STPA step 2)**	**SM-1**	**SM-2**
Feedback and inputs	**RSR-1:** Vehicle sensors must take accurate on time measures on the feedback necessary to determine that ADS is available and that a takeover request is needed	X	X
	RSR-2: Automation must detect when vehicle sensors are providing inaccurate measures with delays regarding the feedback necessary to determine that ADS is available and that a takeover request is needed	X	
	RSR-6: External information (networks) must provide adequate information regarding the feedback necessary to determine that ADS is available and that a takeover request is needed		X
Model	**RSR-7:** Automation must have an adequate model on the ADS conditions indicating that automation can continue on automated driving mode.	X	X
Decision-making	**RSR-9:** Automation's control algorithm must not generate ADS availability notification when the model indicates ADS is not available, and must generate takeover requests when the ADS conditions are no longer met	X	
Action execution	**RSR-11:** Automation must ensure that the actions generated by the control algorithm related to the feedback on ADS is available notification and on takeover requests are sent to the HMI with a maximal delay of TBD.	X	

The highway pilot system takes over lateral and longitudinal control of the vehicle, the object and event detection and response, and the dynamic driving task fallback performance, and therefore, significantly modifies the driving task both for the human driver and the vehicle. Consequently, almost the totality of safety requirements and refined safety requirements for the human driver and automation were allocated to the first mechanism. On the contrary, the second mechanism does not control the vehicle via vehicle actuators; the effects of the second

mechanism are limited to the feedback information provided to automation by the physical and digital infrastructure. As a result, only the categories related to automation were considered in the second mechanism and fewer safety requirements were assigned to the second mechanism.

The second step was to define questions using the safety requirements and refined safety requirements. The questions derived from safety requirements are broader (i.e. they are higher-level requirements) than the ones defined with refined safety requirements; they target the control actions provided by the human driver and automation. The refined safety requirements further "refine" the safety requirements identified using the unsafe control actions into more detailed requirements that target the causes behind unsafe control actions provided by the human driver and automation.

For example, the three safety requirements in table 20 (SR-3, SR-16 and SR-17) were used to define the following question for the evaluation of the first mechanism: "Does automation send adequate feedback to the driver regarding ADS availability notifications and takeover requests?" This very broad question addresses the control actions in which automation sends feedback to the driver to influence a transition. If the analysis needs to be more detailed, the questions related to the reasons why automation may not send adequate feedback can be found using the refined safety requirements. Accordingly, the question derived from the refined safety requirement nine can be examined: "Does automation's control algorithm generate signals to send feedback on ADS availability notification and on takeover requests under adequate contexts?".

These two steps were applied to the safety requirements and refined safety requirements contained in the six categories of unsafe control actions. The results of this process were organized into two tables (table 21 and table 22) which include the category being analyzed, the safety requirements concerned by the questions derived from the safety requirements, and the refined safety requirements concerned by the questions derived from refined safety requirements. The objective of these questions is to provide assistance in the evaluation of direct mechanisms (1-2) which in turn facilitate the assessment of vehicle system's effectiveness.

Table 21 - Questions to consider in the evaluation of the first safety mechanism

The questions are organized according to the 6 categories of unsafe control actions (blue designates that the category is related to automation and green is related to the human driver). The safety requirements used to define general questions and the refined safety requirements used to define detailed questions are indicated.

Category	Safety rqts	General question based on the Safety Requirement	Refined safety rqts	Detailed question based on the Refined safety Requirement
1	SR-3, SR-16, SR-17	Does automation send adequate feedback to the driver regarding ADS availability notifications and takeover requests to influence a transition? *Rationale: The feedback sent by automation to the driver modifies the driving task because it encourages the driver to validate ADS engagement and thus trigger a transition to AD mode, and to validate takeover requests and thus trigger a transition to MD mode.*	RSR-1, RSR-2, RSR-3, RSR-4, RSR-5	Does automation receive adequate information from vehicle sensors on ADS availability and on the need for takeover request?
			RSR-7, RSR-8	Is automation aware when the ADS is unavailable and when takeover requests are needed?
			RSR-9	Does automation's control algorithm generate signals to send feedback on ADS availability notifications and on takeover request under adequate contexts (i.e. when ADS is available and when a takeover request is needed)?
			RSR-10, RSR-11	Do the signals on ADS notifications and on takeover requests reach the HMI with a maximal delay of TBD?
2	SR-4, SR-18, SR-19	Does the driver provide an adequate response to the feedback sent by automation regarding ADS availability notifications and takeover requests? *Rationale: The driver's response to feedback sent by automation modifies the driving task because it triggers a transition to AD mode (after driver's validation of ADS engagement) and a transition to MD mode (after driver's validation of a takeover request).*	RSR-12	Does the HMI display the feedback on ADS availability notifications and on takeover requests?
			RSR-13, RSR-14, RSR-15, RSR-16, RSR-17	Does the human driver perceive feedback displayed by the HMI and feedback on the traffic environment via human perception?
			RSR-18, RSR-19, RSR-20, RSR-21, RSR-22	Is the human driver aware of the HMI feedback, and the feedback on the traffic environment? Does the human driver know how to respond to feedback on ADS availability notifications and on takeover requests?
			RSR-23	Does the human driver have an adequate decision-making process regarding responses to the HMI feedback on ADS availability notifications and takeover requests?
			RSR-24, RSR-26	Do the design of the cockpit and validation procedures support the driver to provide adequate responses to HMI feedback on ADS availability notifications and on takeover requests?
			RSR-25	Is the driver aware of the command location, sequences, etc., necessary to respond to HMI on ADS availability notifications and on takeover requests?
			RSR-27	Do the signals from driver's responses to feedback on ADS availability notification and on takeover requests, reach automation with a maximal delay of TBD?
3	SR-5, SR-6, SR-7, SR-20, SR-21, SR-29, SR-30	Does automation engage and disengage the ADS in appropriate contexts? *Rationale: The control actions provided by automation to engage/disengage the ADS modify the driving task because they determine the transition to AD mode and the start of vehicle control execution by automation, and the transition to MD mode and the end of vehicle control execution by automation.*	RSR-28, RSR-29, RSR-30, RSR-31, RSR-32, RSR-33, RSR-34, RSR-35	Does automation receive adequate feedback from vehicle sensors on ADS engagement/disengagement status, on driver's ADS engagement/disengagement actions, and on ADS conditions?
			RSR-36, RSR-37, RSR-38	Is automation aware of the ADS status, driver's actions ADS engagement/ disengagement actions and ADS conditions?
			RSR-39	Does the control algorithm generate ADS engagement/disengagement under adequate contexts?
			RSR-40, RSR-41	Is the ADS engaged/disengaged when the control algorithm sends the engagement/disengagement control actions?

Category	Safety rqts	Question based on the Safety Requirement	Refined safety rqt	Question based on the Refined safety Requirement
4	SR-12	Does the driver disengage ADS at his/her own request in appropriate contexts? *Rationale: Driver's disengagement at his/her own request (i.e. when there is no takeover request from automation) modifies the driving task because it determines the transition from AD to MD mode and the end of vehicle control execution by automation.*	RSR-42	Does the HMI display correct feedback on ADS status?
			RSR-43	Does the driver perceive the driving environment before ADS disengagement?
			RSR-44, RSR-45	Is the driver aware of ADS status, the procedure to disengage ADS, and the driving environment?
			RSR-46	Does the human driver have an adequate decision-making process regarding ADS disengagements initiated by the driver?
			RSR-47	Does the human driver provide unintended ADS disengagements?
			RSR-48	Do the design of the cockpit and validation procedures limit unintended ADS disengagement?
			RSR-49	Does the ADS disengagement signal reach automation?
5	SR-2, SR-8, SR-9, SR-13, SR-14, SR-15, SR-22, SR-31, SR-25, SR-26, SR-27	Does automation provide adequate control of the vehicle? *Rationale: The execution of vehicle control by automation modifies the driving task.*	RSR-50, RSR-51, RSR-52, RSR-53, RSR-54, RSR-55, RSR-56	Does automation receive adequate feedback on driver's actions, driving mode status, driving environment and ADS conditions?
			RSR-58, RSR-59, RSR-60, RSR-61	Does automation have adequate representations on the driver, ADS status, driving environment, ADS conditions, traffic rules and social norms?
			RSR-62, RSR-63, RSR-64, RSR-65	Does the control algorithm generate adequate control actions for: vehicle control, release of vehicle control, compliance of traffic rules and social norms, and minimal risk maneuver?
			RSR-66, RSR-67, RSR-68	Does the implementation of control actions via vehicle actuators enable adequate vehicle control, release of the vehicle control, compliance of traffic rules and social norms, minimal risk maneuver?
6	SR-1, SR-23, SR-24, SR-28, SR-32, SR-10, SR-11	Does the driver provide an adequate control of the vehicle? *Rationale: The driver's response to feedback sent by automation modifies the driving task because it validates ADS engagement (transition to AD mode) and validates takeover requests (transition to MD mode)*	RSR-69	Does the HMI display correct feedback on ADS status, takeover requests and minimal risk maneuvers?
			RSR-70, RSR-71, RSR-72, RSR-74	Does the human driver perceive feedback displayed by the HMI and feedback on the driving environment?
			RSR-76	Is the human driver aware of the procedures to engage/disengage the ADS and to validate takeover requests?
			RSR-77	Do the driver procedures for ADS operation support a safe operation?
			RSR-78	Does the feedback provided by the HMI assist the driver to safely operate the ADS?
			RSR-79	Does the human driver have an adequate decision-making process regarding vehicle control?
			RSR-82	Is the driver aware of the command location, sequences, etc., necessary to safely operate ADS?
			RSR-83	Do the driver procedures for ADS operation support a safe operation?
			RSR-84	Do the actions provided by the driver reach automation and the vehicle?

Table 22 - Questions to consider in the evaluation of the second safety mechanism

The questions are organized according to the three categories related to automation displayed in blue). The safety requirements used to define general questions and the refined safety requirements used to define detailed questions are indicated.

Category	Safety rqts	General question based on the Safety Requirement	Refined safety rqt	Detailed question based on the Refined safety Requirement
1	SR-3, SR-16, SR-17	Does automation receive adequate feedback from physical and digital infrastructure regarding ADS availability and need for takeover requests? *Rationale: The feedback received by automation on physical and digital infrastructure (along with other feedback)[9] is used by automation to determine if the ADS is available and if a takeover request is needed.*	RSR-1, RSR-5, RSR-6	Does automation measure adequate feedback on physical infrastructure via vehicle sensors and receive adequate information from digital infrastructure (e.g. networks) regarding the ADS availability conditions and the need for takeover requests?
			RSR-7, RSR-8	Does the information measured on physical infrastructure and received via digital infrastructure, enable automation to be aware when ADS is available and when a takeover request is needed?
3	SR-6	Does the feedback on physical and digital infrastructure enable automation to adequately engage/disengage the ADS? *Rationale: The feedback received by automation via networks on ADS conditions (along with other feedback) is used by automation to determine the contexts to engage/disengage the ADS.*	RSR-30, RSR-34, RSR-35	Does automation measure adequate feedback on physical infrastructure via vehicle sensors and receive adequate information from digital infrastructure regarding ADS conditions that affect ADS be engagement/disengagement?
			RSR-37	Does the information measured on physical infrastructure and received via digital infrastructure, enable automation to be aware when ADS conditions affect ADS engagement/disengagement?
5	SR-9, SR-13, SR - 26	Does the feedback on physical and digital infrastructure enable automation to provide adequate control of the vehicle? *Rationale: The feedback received by automation on physical and digital infrastructure (along with other feedback) is used by automation to provide control of the vehicle*	RSR-52, RSR-56, RSR-57	Does automation measure adequate feedback on physical infrastructure via vehicle sensors and receive adequate information from digital infrastructure regarding the driving environment and ADS conditions that influence vehicle control?
			RSR-60,	Does the information measured on physical infrastructure and received via digital infrastructure, enable automation to be aware of the driving environment and ADS conditions that influence vehicle control?

[9] The feedback on physical and digital infrastructure is a part of all the types of feedback that automation receives. Automation also receives feedback on the vehicle, other road users, the human driver, etc.

3.5 Discussion

The discussion of this chapter is organized according to three topics: the target population, the STPA analysis and the resulting safety requirements, and the questions based on safety requirements to assist the evaluation of direct mechanisms.

3.5.1 Target population

(Herve and Lesire 2017) used another French crash database called VOIESUR (for the year 2011) to estimate the target population for an automated driving system similar to the highway pilot system considered in this chapter; VOIESUR is a crash database based on in-depth analyses of police collision reports performed by crash accident experts (Herve and Lesire 2017) found that the automated driving system could potentially address 6% of crash fatalities, 5% of road users injured and hospitalized and 10% of road users injured and not hospitalized. These results are slightly higher than the target population estimated in this chapter using the BAAC crash database: 3,8% of crash fatalities, 3,3% of road users injured and hospitalized and 6,3% of road user injured and not hospitalized. This could be explained by the fact that the estimates were calculated using two databases and data from different years (2011 for VOIESUR and 2015 for the BAAC). Further, there were some differences in the crash variables selected to query the databases, for instance the estimates calculated using the BAAC database omitted crashes involving heavy rain, snow and hail. However, even if the higher target population estimates are chosen, with a fleet penetration rate of 100% and an effectiveness of 100%, these numbers are still very low to have a significant impact on road safety. Does this mean automated driving will have no considerable impact on road safety?

The low numbers for the target population are tightly related to the operational design domain of the highway pilot, notably to the type of road network on which the system can be operated (i.e. highways and other roads with divided carriageway). In France, the percentage of crashes on highways is rather low, for instance only 8%[10] of fatal crashes occurred on highways in 2015;

[10] Some of the crashes on highway network had to be omitted in the target population estimates because the highway pilot system cannot address crashes at intersections, highway exits, involving heavy rain, etc.

the road network with the higher percentage of fatal crashes in 2015 was departmental roads with 64%[11] (ONISR 2017). Consequently, unless automated driving systems are designed to be operated on departmental roads, the road safety impact of this vehicle automation will remain low. Nonetheless, the roadmap for vehicle automation already foresees automated driving systems that can be operated on departmental roads and urban roads such as commuter vehicle systems which enable automated driving on regular daily trips (work-home). Additionally, the ultimate goal of the roadmap is to develop automated driving systems with unlimited operational design domains (SAE level 5) which can manage all driving conditions and potentially address a high rate of crashes.

Limitations:

The main limitation on the estimation of the target population concerns the data contained in the crash database or more precisely, the variables not available in the crash database. For example, the BAAC database does not contain a crash variable for the speed of the vehicles involved in the crash and therefore the estimates may include crashes outside the automated driving system's operational speed range. As a result, the field of crash accident investigation and analysis must prepare for the challenges brought by crashes involving automated driving in terms of new crash variables.

The crashes which only involve material damages are not documented in the BAAC database and therefore they were not taken into account in the estimates. Additionally, the BAAC database only contains crashes on French roads and therefore the target population estimated in this chapter is limited to France and may differ from the estimates in other countries; in order to estimate the overall target population, the target populations in other regions needs to be calculated. The estimation of target populations based crash data does not consider the side effects of vehicle automation and the crashes introduced by automated driving. Finally, although the estimates appear to be low relative to the total number of crashes and injuries

[11] A small portion of the departmental road network has divided carriageway and therefore the part of the crashes on departmental roads with divided carriageway were considered in the target population of the highway pilot system

(partly because all crashes are not reported), the target population will increase once automated driving systems for other types of road networks are developed.

3.5.2 STPA and Safety Requirements

STPA usage:

Although the STPA analysis followed the four parts of the method as described in (Leveson and Thomas 2013), two modifications were made in order to enhance the analysis, namely the elaboration of a graphical timeline covering the control actions and unsafe control actions of the system, and the classification of the entire set of unsafe control actions into six categories. Moreover, the results of the analysis were illustrated using color-coding and a graphical depiction of the control structure and relevant parts of the control loop.

The graphical timeline developed in this chapter demonstrated that diagrams facilitate the representation of the distribution of control actions and unsafe control actions across the phases of the highway pilot system's operation. The timeline provided a comprehensive overview of the entire set of interactions between the human driver and automation which helped to understand the system and group similar unsafe control actions into six categories. Moreover, these depictions served as a communication tool to show others my representation and understanding of the system and as a means to ensure that all the main interactions were considered; experts on the system who looked at the timeline could easily point out when a control action (and by extension unsafe control actions) was missing.

Organizing the 32 unsafe control actions identified in STPA step 1, into six categories, allowed to reduce the processing time of the analysis for the elaboration of scenarios and the number of potential refined safety requirements; instead of analyzing 32 unsafe control actions to generate scenarios, only the 6 categories were examined. Lastly, illustrating the synthesis of the STPA results using color-coding associated to the high-level classes of control flaws (feedback and inputs, model, decision-making and action execution), and displaying the relevant loops on the control structure (as observed in tables 18 and 19), helped novice STPA users to understand the analysis and the interactions being considered. Overall, the use of graphics and illustrations is recommended not only to spark the interest of people unfamiliar with STPA (who are not

always enthusiastic to look at huge tables), but also to structure the results of the analysis in an organized and synthetized fashion.

Safety requirements:

One of the key contributions of the STPA analysis was the identification of unsafe control actions and corresponding safety requirements related to all the phases of the highway pilot 's operation, not only of those related to the automated driving phase (while the system is engaged) and the transition phase from automated driving to manual driving. Conducting a hazard analysis which directly starts by the assumed hazardous and unsafe interactions can lead to focus the analysis on the "critical" phases such as automation does not perceive another road user during automated driving and the human driver does not respond to the takeover request sent by automation. Instead, the STPA analysis begins by considering all the interactions then examines whether or not they are unsafe in the STPA step 1. Accordingly, in addition to unsafe control actions during automated driving and transitions from automated driving to manual driving, the STPA analysis identified several unsafe control actions during the manual driving phase (before the engagement of the system) such as automation provides control of the vehicle during manual driving or automation sends ADS availability notification when the ADS is not available. Furthermore, the STPA analysis also identified unsafe control actions during the transition from manual driving to automated driving and the manual driving after the disengagement of the system.

Moreover, the STPA step 2 (which further examines the reasons behind unsafe control actions) showed that the scenarios leading to the identified unsafe control actions associated to the human driver and automation involve flaws related to feedback, process model, the decision-making process and the execution of actions. Nonetheless, most of scenarios included feedback flaws and process model flaws. For example, automation provides inadequate control of the vehicle during automated driving because it receives inadequate feedback on the driving environment. Also, automated driving does not send a takeover request when the ADS conditions are no longer met because automation is not aware that ADS conditions are not satisfied. Therefore, the elaborated scenarios indicate that the evaluation of assumptions on

feedback and on process models is essential for the safety benefit assessment of automated driving systems.

Limitations:

The classification of the 32 unsafe control actions into six categories may eliminate some of the specificities related to the individual unsafe control. However, these categories were necessary to reduce the processing time of the elaboration of scenarios and definition of refined safety requirements. The results obtained with the classification provided scenarios with a sufficient level of detail for the aim of the analysis (i.e. providing requirements that enable the definition of questions related to the evaluation of direct safety mechanism); if necessary, the persons conducting the safety benefit assessment can look at the unsafe control actions contained in every category and decide to examine them individually in order to generate more detailed scenarios and thus more detailed questions.

A second limitation concerns the ability of the highway pilot system control structure to accurately represent the real system; an analysis on an inconsistent model of the highway pilot system would lead to incorrect and incomplete results. Although multiple documents and meetings with experts on the system were considered to build the control structure, the control structure may differ from the real system. Nevertheless, this issue was partly addressed by having two system experts review, modify and validate the control structure. Further, the results of the analysis are also affected by the experience and system understanding of the person(s) conducting the analysis. The person who conducted the analysis was not an expert on the system; however, the results of the analysis were discussed with experts on the system in order to validate them.

3.5.3 Questions derived from the safety requirements

The findings of this chapter illustrate how the safety requirements and refined safety requirements of the STPA analysis can be used to define questions that aim to address the assumptions related to the evaluation of the direct safety mechanisms (i.e. safety mechanisms 1-2). The questions are organized according to the results of the STPA analysis and therefore

they are divided into the six categories of unsafe control actions which contain questions regarding flaws in feedback, process models, decision-making processes and action execution.

Although it is necessary to test the questions derived from safety requirements with empirical data to analyze if they facilitate the evaluation of the direct safety mechanisms; the questions based on the safety requirements can be compared with broad questions generated without a systematic method in order to observe the usefulness of using structured analysis to define questions. For example, table 23 compares a broad question regarding the impact of takeover requests on the driving task with the general questions and detailed questions derived from safety requirements for the safety mechanism one. As seen in the table, the questions derived from the STPA analysis address a variety of specific factors regarding the takeover request; on the one hand, there are questions related to the takeover notification that automation sends to the driver such as the feedback from vehicle sensors, automation's model on the need of a takeover requests, etc. On the other hand, there are questions associated to the driver's response to the takeover request including the perception of the feedback, the knowledge and driver's models on takeover request procedures, the design of the system, etc.

Table 23 – Comparison of broad question and questions derived from the STPA analysis

Broad question	**Questions derived from the STPA analysis**	
	General questions	**Detailed questions**
What are the direct impacts of takeover request on the driving task?	Does automation send adequate feedback to the driver regarding ADS availability notifications and takeover requests to influence a transition?	Does automation receive adequate information from vehicle sensors on ADS availability and on the need for takeover request?
		Is automation aware when the ADS is unavailable and when takeover requests are needed?
		Does automation's control algorithm generate signals to send feedback on ADS availability notifications and takeover request under adequate contexts (i.e. when ADS is available and when a takeover request is needed)
		Do the signals on ADS notifications and takeover requests reach the HMI with a maximal delay of TBD?
		Does the HMI display the feedback on ADS availability notifications and takeover requests?
	Does the driver provide an adequate response to the feedback sent by automation regarding ADS availability notifications and takeover requests to influence a transition?	Does the human driver perceive feedback displayed by the HMI and feedback on the traffic environment via human perception?
		Is the human driver aware of the HMI feedback, and the feedback on the traffic environment? Does the human driver know how to respond to feedback on ADS availability notifications and takeover requests?
		Does the human driver have an adequate decision-making process regarding responses to the HMI feedback on ADS availability notifications and takeover requests?
		Do the design of the cockpit and validation procedures support the driver to provide adequate responses to HMI feedback on ADS availability notifications and on takeover requests?
		Is the driver aware of the command location, sequences, etc., necessary to respond to HMI on ADS availability notifications and on takeover requests?
		Do the signals from driver's responses to feedback on ADS availability notification and on takeover requests, reach automation with a maximal delay of TBD?

Limitations:

The main limitation of the questions derived from safety requirements is that they have not been applied on a safety benefit assessment; as a first step, they provide assistance to evaluate the assumptions related to direct safety mechanisms but their usefulness on a real safety benefit assessment needs to be addressed. Moreover, the most relevant means to examine the questions also need to be defined, which question need to be examined with questionnaires, with studies on driving simulators, on closed roads or on open roads?

Finally, the research presented in this chapter only looked at the direct mechanism (1-2); the other safety mechanisms (3-9) must also be considered for a comprehensive assessment of the effect of automated driving systems. While STPA analyses have the potential to address the indirect safety mechanisms (3-5) related to the risk dimension of road safety, the mechanisms related to exposure and accident consequences demand other types of methods (e.g. traffic counts, questionnaires and interviews, in-depth accident analyses) and are outside of the scope of all hazard analyses methods.

3.6 Conclusions

This chapter introduced an approach to contribute to the broader process of the safety benefit assessment of automated driving systems by estimating the target population and by defining questions based on STPA safety requirements, to assist the evaluation of direct safety mechanisms. While the target population estimates show that the highway pilot system has a limited potential effect on road safety (it addresses less than 5% of injury crashes and less than 4% of fatalities), the roadmap to automated driving systems forecasts systems which can be operated in other types of road network with higher rates of crashes and fatalities, and therefore higher potential for road safety improvements.

Concerning the STPA methodology, the classification of the 32 unsafe control actions into 6 categories enabled to reduce the analysis time of scenarios and refined safety requirements (i.e. STPA step 2). Also, the use of color coding and graphical depictions such as the timeline, the synthesis of STPA results, etc. facilitated the STPA analysis and the communication of the STPA results.

Finally, the results of the STPA analysis demonstrated that the assumptions related to the evaluation of direct safety mechanisms (e.g. the proper functioning and safe operation of an automated driving system, and the safe interactions between the human driver and automation) are not always explicitly stated and may require a comprehensive analysis. The safety requirements and refined safety requirements identified though the STPA analysis allow to derive questions related to those assumptions which can be subsequently tested in studies on driving simulators, on closed and semi-private roads and also field operational trials on open roads, to further examine the evaluation of direct safety mechanisms.

3.6.1 Future work

The approach described in this chapter to estimate the target population of automated driving systems and to assist the evaluation of direct safety mechanisms, should be applied on other systems with less limited operational design domains, in order to examine the full potential of vehicle automation.

Renault participates in a European Project called L3Pilot which aims at testing the viability of automated driving as a safe and efficient mode of transportation and evaluate the expected benefits of ADS (including safety benefits) using field FOTs in 11 European Countries. As a part of the project, Renault will test several prototypes of a highway pilot system on open roads. L3Pilot provides a unique opportunity to investigate the usefulness of the questions defined using the safety requirements; these questions can help the designers of the FOT determine what they need to evaluate regarding safety benefit assessment, in the trials. Furthermore, the data collected in the trial will provide evidence on the relevance of the questions for the assessment of direct safety mechanisms. Finally, the application of STPA to derive questions related to indirect safety mechanisms (3-5) and the integration of this approach with the methods and results related to the other safety mechanisms (6-9) should also be investigated.

Résumé chapitre 4: Sécurisation des expérimentations des véhicules autonomes

La seconde question de recherche « Comment sécuriser les expérimentations des véhicules autonomes » est traitée dans ce chapitre par la constitution d'un cadre des exigences de sécurité sur les expérimentations des véhicules autonomes. Deux analyses STPA ont été menées pour définir les contraintes de sécurité sur le système. La première analyse se concentre sur le système français des expérimentations du véhicule autonome au niveau macroscopique (gouvernement, organismes de financement, constructeurs automobile à tous les niveaux). La seconde analyse, faite à un niveau plus microscopique, se penche sur l'expérimentation du véhicule autonome menée par Renault sur un système « highway pilot system ». Le cadre des exigences de sécurité est donc constitué des contraintes de sécurité provenant des deux analyses, sections une à quatre pour l'analyse macroscopique et section cinq pour l'analyse microscopique. Enfin, les résultats du chapitre sont discutés en rapport avec le périmètre et le contenu du cadre des exigences et avec les limitations de cette approche.

Chapter 4: Using STPA to ensure the safety of automated driving trials

4.1 Chapter overview

This chapter introduces a framework to ensure the safety of the entire automated driving trial process. As illustrated in figure 30, two STPA analyses are conducted to define safety requirements on the system's behaviors. The first analysis is performed on the French vehicle trial process and the second analysis on an automated driving trial involving a highway pilot system conducted by Renault. The safety requirements resulting from the first analysis are organized to create sections 1-4 of the framework. Additionally, the safety requirements identified through the second analysis are organized to create the fifth section of the framework. Lastly, the findings of the chapter are discussed relative to the scope and contents of the framework and the limitations of the approach.

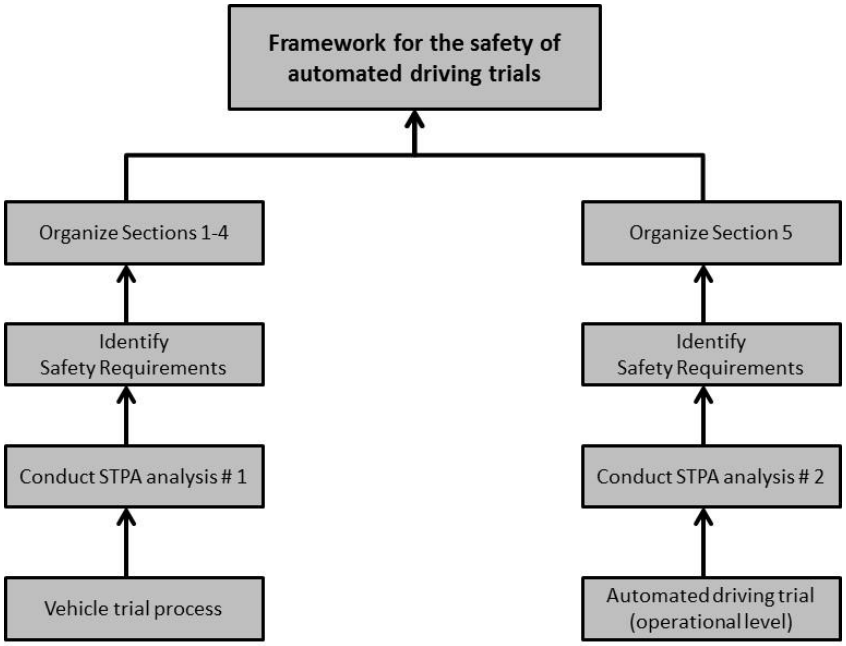


Figure 30 - Process to establish the framework to ensure the safety of automated driving trials

4.2 Introduction

Vehicle automation is expected to bring benefits such as improving road safety, reducing traffic congestion, decreasing emissions and energy consumption, mobility for everyone, free time and more comfort, etc. (Maurer et al. 2016). Consequently, the governments of several countries are supporting and facilitating the development of automated driving technologies. However, the organizations developing automated driving technologies must conduct vehicle trials to validate their technical performance reliability, their acceptability, and to assess their impacts, before their market introduction. While the initial phases of automated driving trials can be done in test laboratories and controlled environments via computer simulation, component testing, driving simulators, and tests tracks; vehicle trials on public roads are necessary in order to test automated driving technologies in real-world settings.

As large-scale automated driving trials on open roads emerge in several countries, the government, automakers and the other stakeholders involved in the trials, need to address the question of how to ensure the safety of such trials (Hottentot, Meines, and Pinkaers 2015; UK Department of Transport 2015b; "DAVI – Dutch Automated Vehicle Initiative" 2017, "Drive Me" 2017). On the one hand, governments have established procedures to obtain authorizations for automated driving trials in which the company requesting the authorizations needs to demonstrate that sufficient in-house testing has been conducted and that they are ready for trials on public roads . For instance, the French government established a dossier for open road trials which includes a set of conditions that must be satisfied to obtain a trial authorization (Ministère de l'environnement, de l'énergie et de la mer, chargée des relations internationales sur le climat 2015). Moreover, the UK department of transport defined a code of practice to provide guidance for organizations that intend to conduct automated driving trials (UK Department of Transport 2015a). Also, the Society of Automotive Engineers (SAE) published guidelines of the safe conduct of on-road tests of automated driving system prototypes SAE levels 3-5 (SAE International 2015). On the other hand, automakers and other organizations conducting vehicle trials also establish internal standards and procedures to ensure the safety of automated driving trials which supplement the minimal requirements defined by the government.

Nevertheless, there is no safety initiative that considers the concerns and actions of the government and vehicle manufacturers, as well as the interactions across all the levels of the entire vehicle trial sociotechnical system. For example, the influence of the company management on the development of hazardous vehicle prototypes and hazardous vehicle processes is not taken into consideration in the trial safety initiatives such as guidelines and company standards. Additionally, the safety initiatives are being established by employing traditional risk analysis approaches (e.g. Preliminary Risk Analysis, FMEA, etc.) that may not fully capture the hazards introduced by automation such as software flawed requirements, human unsafe interaction with automation, and unsafe system behaviors in which no component failure is involved (Leveson 2004, 2011). Therefore, there is a need for a common hazard analysis approach capable of considering the entire sociotechnical system and the hazards introduced by vehicle automation, to ensure the safety of automated driving trials.

In the aviation field (Montes 2016) has already explored the application of STPA on safety planning of a real flight test project in which an autonomous wingman system enabled an unmanned aircraft to fly in formation with respect to a lead manned aircraft. He developed an STPA-based framework for the elaboration of safety plans which analyzed the safety of the flight test procedure and the inherent safety of the autonomous wingman system. Furthermore, the comparison of the safety plan elaborated with STPA and the traditional safety plan established by the test project organizers (Montes did not participate in the elaboration of the traditional safety plan) demonstrated that STPA identified more minimizing procedures, corrective actions and recovery actions than the traditional safety plan. Consequently, the application of STPA to trial safety involving automated systems, and their use on automated driving trials should be further investigated.

4.2.1 Study aim and objectives

The aim of this chapter was to tackle the third research question *"how to ensure the safety of automated driving trials?"* by examining how STPA can contribute to ensure the safety of automated driving trial processes.

The following objectives were defined to achieve the aim of the study:

- Analyze the vehicle trial sociotechnical system using an STPA analysis to identify the safety requirements needed to ensure the safety of the entire vehicle trial process.
- Analyze a real driving trial involving a highway pilot system using an STPA analysis to identify the safety requirements needed to ensure the safety of an automated driving trial operation process.
- Build a framework based on the classification of the outputs of the two analyses to ensure the safety of automated driving trial processes and to identify actions to be implemented by the car manufacturers to conduct safe trials.

4.3 Methods

This section discusses the methods employed to conduct the two[12] STPA analyses and to classify their resulting safety requirements into a five section framework to ensure the safety of automated driving trials.. The first STPA analysis intended to capture the hazards and safety requirements enforced at the higher-levels of the vehicle trial process (e.g. the government, company management, trial manager, etc.) in order to structure sections 1-4 of the framework.

The second STPA analysis aimed at capturing the hazards and safety requirements at the operational level of a specific vehicle trial involving a highway pilot system, to establish the fifth section of the framework. Although some of the results of the second analysis are applicable to all automated driving trials, the hazards and safety requirements at the operational level depend on the conditions and specificities of every trial. For instance, a vehicle trial with a supervisor and trained driver drivers will not have the same hazards as a trial with no supervisor and a novice driver.

[12] It was decided to conduct two separate STPA analyses because the higher levels of the system do not change drastically depending on trial operations. On the contrary, the conditions and different vehicle systems tested at the lowest level are subjected to more variations. It is assumed that future analyses will be conducted at the lowest level without redoing the analysis at the higher levels of the system.

4.3.1 STPA analysis on the vehicle trial process

The STPA analysis was conducted on Renault's vehicle trial process for open road testing in France. The data for the analysis were collected from two company standards on vehicle trials and a semi-constructive interview with the company employee in charge of automated driving trials.

The STPA analysis on the vehicle trial was performed according to four stages:

1. Definition of the system engineering foundation: the system engineering foundation for the analysis was established by defining the accidents, hazards and constraints at the system level systems, and by building the control structure for the vehicle process. Moreover, the control structure was validated with the company employee in charge of preparing automated driving trials.
2. Identification of unsafe control actions (STPA step 1): the unsafe control actions were identified by analyzing the control actions of the control structure relative to the second type of unsafe control actions i.e. an unsafe control action is provided that leads to a hazard.
3. Definition of safety requirements: the identified unsafe control actions were translated into safety requirements.
4. Elaboration of scenarios leading to unsafe control actions (STPA step 2) and definition of refined safety requirements: The mental model flaws and feedback flaws were used to generate scenarios leading to unsafe control actions and to define refined safety requirements.

4.3.2 STPA analysis on an automated driving trial operation

The second STPA analysis was performed on the operational level of an automated driving trial involving a highway pilot system. The data for analysis were collected from discussions held during the trial design meetings and from trial design documents.

The STPA analysis was performed following four parts:

1. Definition of the system engineering foundation: since the losses for the operation level are the same as the losses for the vehicle trial process, the accidents, hazards and

constraints at the system level were the ones defined in the first STPA analysis. The control structure for the trial involving the highway pilot system was built using the data from the design meetings and the trial design documents.

2. Identification of unsafe control actions (STPA step 1): unsafe control actions were identified by analyzing the control actions of the control structures relative to the first two types of unsafe actions i.e. a control action required for safety is not provided, and an unsafe control action is provided that leads to a hazard.
3. Definition of safety requirements: the identified unsafe control actions were translated into safety requirements.
4. Elaboration of scenarios leading to unsafe control actions (STPA step 2) and definition of refined safety requirements: The control flaws classification was used to generate scenarios leading to unsafe control actions and to define refined safety requirements.

4.3.3 Framework to ensure the safety of automated driving trials

The framework was created by classifying the safety requirements defined via the two STPA analyses into five sections. The safety requirements identified through the first STPA analysis on the vehicle trial process were classified into sections 1-4; and the safety requirements identified through the second STPA analysis on an automated driving trial operation were classified into section 5.

Sections 1-4 of the framework

The responsibilities of the controllers involved in the vehicle trial process were examined to identify four joint-responsibilities which constitute the sections 1-4 of the framework:

- **Section 1:** Definition of policies and resources for vehicle technology development and vehicle trials.
- **Section 2:** Establishing orientations for vehicle technology development and vehicle trials.
- **Section 3:** Approval of the trial.
- **Section 4:** Design and development of vehicle trial.

Additionally, the fourth section was divided into three sub-sections:

- **Sub-section 4.1:** Organization and preparation of the trial.
- **Sub-section 4.2**: Trial data.
- **Sub-section 4.3:** Safety and compliance of the trial.

Further, clusters within each section were created to group the safety requirements defined through the first STPA analysis. Lastly, the clusters were used to define categories of safety requirements for sections 1-4.

Section 5 of the framework

The main responsibility of the vehicle trial operation system regarding safety which is to ensure the safety of the trial operation, was used to establish section 5 of the framework.

- **Section 5:** Safety of trial operation.

Moreover, section 5 was further divided into three sub-sections:

- **Sub-section 5.1:** Safety related to the maturity level of the vehicle technology being tested.
- **Sub-section 5.2:** Safety related to the vehicle trial.
- **Sub-section 5.3:** Trial operation data.

As in sections 1-4, clusters were created to organize the safety requirements identified in the second STPA analysis and to define categories of safety requirements for section 5.

4.4 Findings

Section 4.4 presents the outputs of the two STPA analyses and the classification of the identified safety requirements into a framework to ensure the safety of automated driving trials.

4.4.1 STPA analysis on the vehicle trial process

The STPA analysis involved establishing the system engineering foundation for the analysis, the identification of unsafe control actions, the definition of safety requirements, the elaboration of scenarios leading to unsafe control actions and the definition of refined safety requirements.

System engineering foundation for the analysis

The system engineering foundation established for the analysis comprised the definition of two system accidents, two hazards, two safety constraints, and the construction of the vehicle trial control structure which includes the government, funding agencies, several actors within Renault and the vehicle trial operation process.

System accidents

ACC-1: People die or get injured during a vehicle trial.

ACC-2: Property damage during a vehicle trial.

System hazards

H-1: The vehicle violates safety distance to other road users or objects on the road during a vehicle trial.

H-2: The vehicle leaves the roadway during a vehicle trial.

System safety constraints

SC-1: The safety control structure must prevent the vehicle from violating safety distance to other road users or objects on the road during a vehicle trial.

SC-2: The safety control structure must prevent the vehicle from leaving the roadway during a vehicle trial.

Control structure of the vehicle trial process

The control structure of the vehicle trial process (displayed in figure 31) models the interactions of stakeholders at several levels of the sociotechnical system. The highest level covers the government who establishes regulations to enforce the safety of vehicle trials such as the W garage certificate (which is a temporary car plate that vehicle prototypes need to have to travel on open roads). Moreover, the government also demands automakers and other organizations conducting automated driving trials, to request a trial authorization (Ministère de l'environnement, de l'énergie et de la mer, chargée des relations internationales sur le climat 2015). The government receives feedback from the lower levels of the system via hearings, meetings, and the dossiers to request certificates and authorizations.

The second highest level includes the agencies that provide funding and requirements for vehicle trials (i.e. requirements to assess the impact on vulnerable road users on specific types of road networks). In turn, the organizations conducting the vehicle trials send feedback to funding agencies through vehicle trial proposals.

The third level comprises seven stakeholders within the vehicle company, which play a role in the vehicle trial process.

1. The company management: they define the roadmap that sets the orientation for future vehicle technology developments and vehicle testing. Also, they establish standards and resources to follow the roadmap. In terms of feedback, the company management receives trial results and change reports from the lower levels of the company.
2. The department authorizing the vehicle trial: they authorize (or not) the trial based on the approval request delivered by the trial manger, the expert recommendations and the trial consent granted by the department providing the prototype.
3. Company expert leaders: they provide recommendations to ensure the compliance and safety of the trial. Additionally, they receive feedback from the trial manager regarding the vehicle trial.
4. The department providing the prototype: they give consent to use the prototype under the vehicle trial conditions. They receive feedback on the vehicle trial from the trial manager and provide information on the vehicle technology prototype to trial executors. Furthermore, they may work with service providers for the development of the prototype.
5. The trial manager: s/he defines the objectives and conditions for the trial, coordinates the trial, assesses safety and compliance, and receives feedback from the trial executor.
6. The trial executor(s): they define the protocol for the trial and the data recording specifications, collaborate with company teams and service providers to meet the trial manager demands and prepare the trial. In terms of feedback, they receive information from the company teams, service providers, the department providing the prototype, and data from the vehicle trial operation process.

7. The company teams and service providers: they implement the requirements and specifications established by the trial executor(s), and inform the trial executor(s) of the problems and changes in trial requirements.

Lastly, the lowest level of the structure contains the vehicle trial operation process with the actors at the sharp end of the system (i.e. the trial staff, trial experimenter, driver participant and automation). The vehicle trial operation is analyzed in the second STPA analysis.

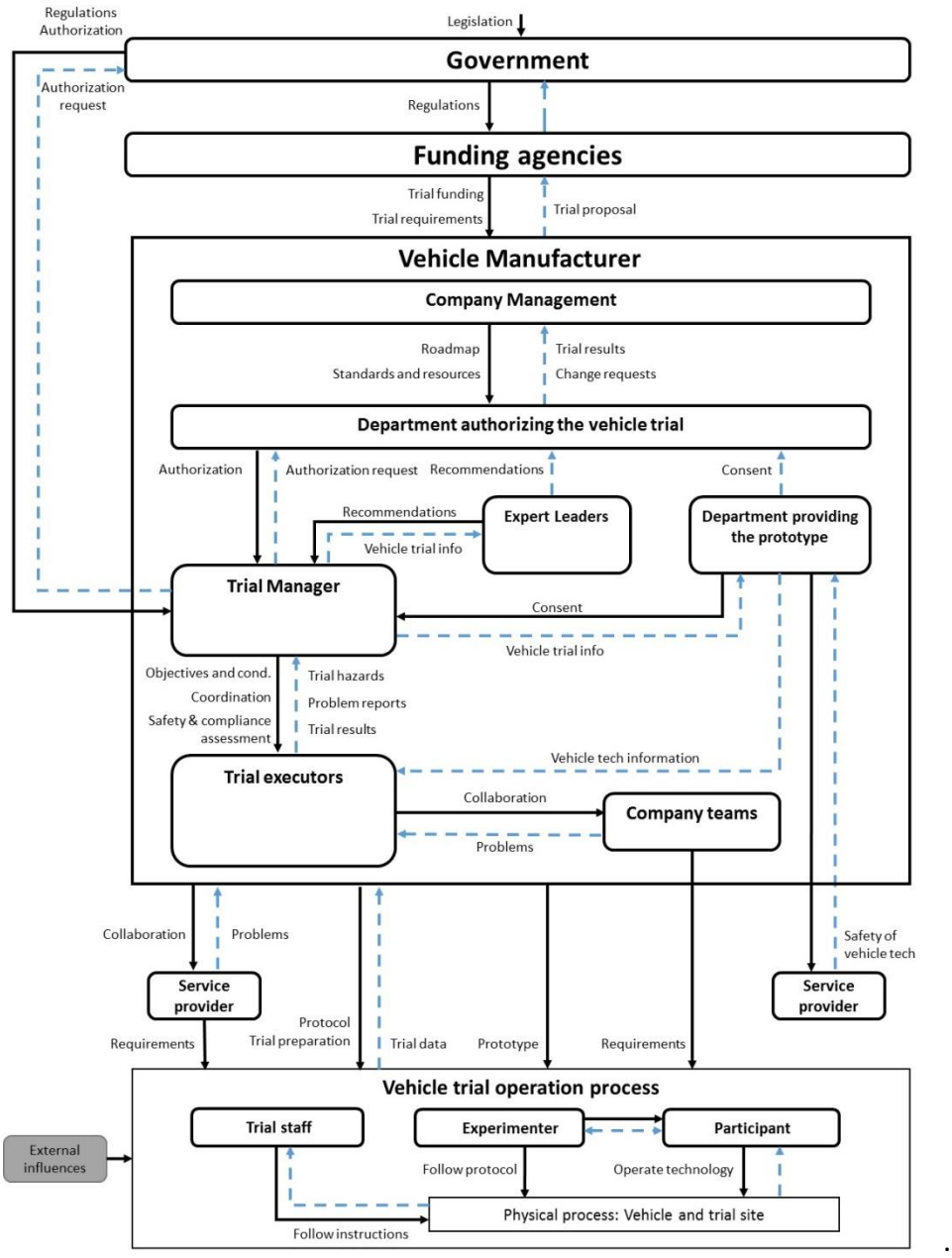


Figure 31 - Control structure of the vehicle trial process

Unsafe control actions (STPA step 1)

The control actions of the control structure shown in figure 31 were examined according to the second type of unsafe control actions (i.e. an unsafe control action is provided that leads to a hazard) to identify 17 unsafe control actions. For organizational and human controllers such as the government, company management and trial managers, the second type of unsafe control actions was enough to identify all unsafe control actions. For example, the unsafe control action in which the government does not establish regulations for vehicle trials or establishes regulations too late, or stops establishing regulations too soon, can be considered as a subset of the unsafe control action in which the government establishes inadequate regulations.

Table 24 displays some examples of the unsafe control actions established in the first STPA analysis (see appendix B for the complete list of unsafe control actions).

Table 24 - Examples of unsafe control actions for the first STPA analysis

Hazards: Violating safety distance and leaving roadway during a vehicle trial		
Controller	**Control action (CA)**	**Providing the CA causes hazard**
Government	Authorize trial	**UCA-2:** The government authorizes an unsafe vehicle trial
Company Management	Define roadmap	**UCA-5:** The company management defines an inadequate roadmap that facilitates the development of unsafe vehicle technologies and unsafe vehicle trials
Trial Manager	Define objectives and conditions	**UCA-10:** The trial manager defines trial objectives and conditions that contribute to an unsafe vehicle trial
		UCA-12: The trial manager inadequately assesses the compliance and the safety of the trial

Safety requirements

The 17 unsafe control actions identified through the first step of the STPA analysis were used to define 17 safety requirements as illustrated in the following examples:

- **UCA-2:** The government authorizes an unsafe vehicle trial.
- **SR-2:** The government must not authorize an unsafe vehicle trial.

- **UCA-5:** The company management defines an inadequate roadmap that leads to the development of unsafe vehicle technologies and unsafe trials.
- **SR-5:** The company management must define an adequate roadmap that facilitates the development of safe vehicle technologies and safe vehicle trials.

- **UCA-10:** The trial manager defines trial objectives and conditions that contribute to an unsafe trial.
- **SR-10:** The trial manager must not define trial objectives and conditions that contribute to an unsafe trial.

- **UCA-12:** The trial manger inadequately assesses trial compliance and safety.
- **SR-12:** The trial manger must adequately assess trial compliance and safety.

Scenarios leading to unsafe control actions and refined safety requirements (STPA step 2)

The control flaws in mental models and feedback loops[13] were used to examine the 17 identified unsafe control actions provided by the high-level controllers of the vehicle trial process, in order to elaborate scenarios leading to the unsafe control actions and to define 41 refined safety requirements. The flaws in the controller's mental models (like inconsistent representation of trial safety or incorrect model of the vehicle technology being tested), and the flaws in the feedback received by higher-level controllers from the lower-levels, were found to be the main reasons why controllers provide unsafe control actions, and allowed the definition of more detailed safety requirements i.e. refined safety requirements.

Table 25 illustrates some examples of the process in which unsafe control actions are examined to elaborate scenarios and define refined safety requirements. For instance, in UCA-2, the government may authorize an unsafe vehicle trial because they are not aware that the trial is unsafe. As a result, there are refined safety requirements on the government's mental model of the vehicle trial and on the feedback regarding the vehicle trial that the government receives from the trial manager via the authorization request.

[13] Several iterations of the analysis were made in which we tried incorporating the other categories of the classification (e.g. inadequate sensor operation, inadequate control algorithm, inappropriate control action, etc.) to generate additional scenarios. However, at this level of abstraction, the other control flaws did not bring more fundamentally different causes for unsafe control actions. Therefore the flaws in mental models and feedback were enough to analyze the reasons why the controllers of the vehicle trial process provide unsafe control actions.

Table 25 - Examples of scenarios and refined safety requirements for the first STPA analysis

Unsafe control action	Scenario	Refined safety requirement
UCA-2: The government authorizes an unsafe vehicle trial	Mental model flaw (trial safety): The government authorizes an unsafe vehicle trial because they are not aware that the trial is unsafe	**RSR-2.1:** The government must have an adequate model of the vehicle trial
		RSR-2.2: The trial manager must provide adequate feedback in the dossier for a trial authorization request
UCA-5: The company management defines inadequate roadmap that facilitates the development of unsafe vehicle technologies and unsafe vehicle trials	Mental model flaw (need for a clear roadmap): The company management defines an inadequate roadmap because they consider that the roadmap does not need to be clear an understandable for all employees	**RSR-5.1:** The company management must define a clear and understandable roadmap and diffuse it to all employees
	Mental model flaw (Roadmap's safety): The company management defines an inadequate roadmap because they have an incorrect model of the roadmap's safety (they believe that it is safe when it is not)	**RSR-5.2:** The company management's model must include the knowledge and information necessary to assess roadmap's safety
		RSR-5.3: The lower levels of the company must provide company management with adequate feedback on hazards associated to vehicle technologies and vehicle trials

The results of the first STPA analysis (i.e. the 17 safety requirements defined based on the 17 identified unsafe control actions and the 41 refined safety requirements) provided the inputs for sections 1-4 of the framework. These requirements address the behavior of the high-level controllers in the vehicle trial process, which in turn, set the context for the vehicle trial operations.

The requirements on regulations, company standards, roadmap for vehicle technology, expert recommendations, trial approval a, design of vehicle trials, etc. influence the safety of all vehicle trial operations. Instead, the second STPA analysis aims to focus on the operation of a specific vehicle trial involving a highway pilot system to identify requirements at the lowest level of the vehicle trial system which provide the inputs for the fifth section of the framework.

4.4.2 STPA analysis on an automated driving trial operation

The STPA analysis was performed by establishing the engineering foundation for the analysis identifying unsafe control actions, defining safety requirements, and elaborating scenarios that lead to unsafe control actions which enable to defining refined safety requirements.

System engineering foundation for the analysis

The system engineering foundations consist of defining the accidents, the hazards and the constraints at the system level and building the control structure of the system being analyzed. Since the losses for the second STPA analysis are the same as the ones in the first STPA analysis, the definitions for accidents, hazards and constraints are the same. On the other hand, the control structure built for the second STPA analysis represents the lowest level of the control structure in figure 31 examined in the first STPA.

System accidents

ACC-1: People die or get injured during a vehicle trial.

ACC-2: Property damage during a vehicle trial.

System hazards

H-1: The vehicle violates safety distance to other road users or objects on the road during a vehicle trial.

H-2: The vehicle leaves the roadway during a vehicle trial.

System safety constraints

SC-1: The vehicle must not violate safety distance to other road users or objects on the road during a vehicle trial.

SC-2: The vehicle must not leave the roadway during a vehicle trial.

Control structure of an automated vehicle trial operation

While the highway pilot system (see section 3.3.1 in chapter 3 for a description of the system) has already been tested at the component-level, on closed test tracks to validate the technical readiness of the prototype, on driving simulators to validate the HMI interfaces, and on open

roads with expert drivers; the automated driving trial operation considered in this study, is the first trial in which lambda or novice drivers (i.e. non-expert drivers or trained test drivers) operate the highway pilot system on open roads.

The main objectives of the trial are: (a) to evaluate driver's acceptability of the highway pilot system; and (b) to evaluate driver's behaviors and understanding of the highway pilot system, particularly during the takeover request.

The safety expert responsible for the vehicle trial conducted a preliminary risk analysis and a Failure Mode and Effect Analysis (FMEA) to establish the main safety measures. Accordingly, the prototype was equipped with a double pedal and double steering system to allow the intervention of a trial supervisor when necessary. Moreover, the trial vehicle is also equipped with an emergency switch that interrupts the power supply for the automated driving system, and a stop button that turns off the engine of the vehicle.

The trial is divided into two phases: a training phase and a driving phase. For the training phase, 15 participants are given information on the highway pilot system, the HMI, the exit modes and takeover sequences. Subsequently, to learn how to use the system and how to respond to takeover requests, participants drive the trial vehicle and operate the highway pilot system on closed test-tracks. During test-track driving a trial supervisor is always present as a co-driver and ready to intervene if necessary. For the driving phase, participants drive the trial vehicle on a 90 minute open-road route previously defined by trial organizers, and operate the highway pilot system when the AD mode is available. The trial supervisor is also present and ready to intervene during open-road driving. Additionally, there is a trial experimenter in the rear seat of the vehicle who observes the trial, asks the participants to verbalize their driving experience and collects data on the trial. Furthermore, data related to the driving behavior such as reaction times for the takeover requests, gaze results, and vehicle trajectory after takeover requests, are also recorded. After the driving phase, participants are interviewed and asked to fill-out a marketability questionnaire.

The control structure built for the highway pilot system trial described above is displayed in figure 32. It considers the controllers at the operational level of a trial such as the trial staff, the

trial experimenter, the driver participant, automation, and the trial supervisor. In terms of interactions with the higher levels of the vehicle trial system, the trial design team provides controllers at the operational level with the training necessary for the trial, the trial configurations defined in the trial design and the instructions and protocol for the trial. Moreover, the data recorded during trial operation are sent to trial design team for data processing and analysis.

At the operational level, the trial staff is in charge of managing trial logistics, securing the trial site and implementing the instructions defined by the trial design team to ensure the safety of the vehicle and the people involved in the trial.

The trial experimenter has several responsibilities:

1. S/he must follow the protocol.
2. Provide the participant with instructions on safety, instructions on how to operate the vehicle technology and instructions on what to do during the trial.
3. Interact with the participant.
4. Ensure data recording.

The driver participant is expected to control the vehicle and to comply with traffic rules during manual driving, to release the control of the vehicle when the automated driving system is engaged and to respond to takeover requesst. On the other hand, automation is expected to propose the AD mode when the mode is available, to control the vehicle when the AD mode is engaged (i.e. during automated driving), to send a takeover request when automation reaches the limits of its operational design domain (end mode types 1 and 2 described in section 3.3.1), and to execute a minimal risk maneuver when the driver does not respond to the takeover request or when the vehicle reaches the limits of its operational design domain and cannot assure safe operation for a few seconds (end mode type 3).

Finally, the trial supervisor has the sole responsibility of intervening when the human driver or automation put the vehicle in an unsafe situation; the trial supervisor can intervene through a set of double commands that override the control of the vehicle, an emergency switch that cuts the power supply of the automated driving system and a stop button that shutdowns the engine.

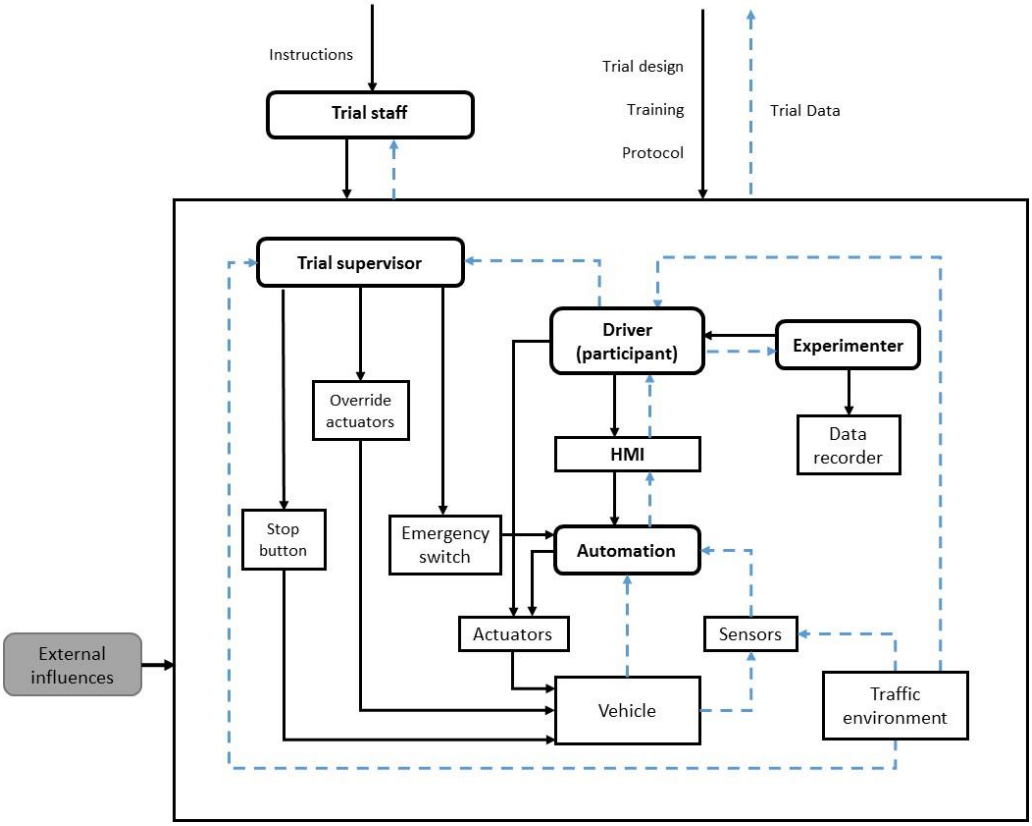


Figure 32 - Control structure of the automated driving trial operation process

Unsafe control actions (STPA step 1)

After the definition of the accidents, the hazards, and the constraints at a system level, and building the control structure of the system, the first step of an STPA analysis was performed to examine the control actions in order to identify unsafe control actions. For this analysis, the first and second types of unsafe control actions (i.e. a control action required for safety is not provided and an unsafe control action is provided that leads to a hazard) were used to examine the control structure displayed in figure 32, and to identify 18 unsafe control action (see appendix C for all the results of the second STPA analysis). The third and fourth types of unsafe control actions related to the order and duration of the control action, are considered as a subset of unsafe control actions in which an unsafe control action is provided that leads to a hazard. For instance, the unsafe control action in which the trial supervisor intervenes too late, is considered as a part of the unsafe control action in which the supervisor intervenes and puts the vehicle in an unsafe situation.

Table 26 illustrates five examples of unsafe control actions identified through the STPA analysis. The UCA-6 considers the context in which the driver provides a control action (provides inadequate control of the vehicle) and causes a hazard. Normally, the driver's control action of providing control of the vehicle during manual driving is a potentially safe control action which is executed to keep the vehicle in safe situations, but when the driver provides inadequate control of the vehicle (e.g. s/he does not apply brakes when the safety distance to other road users is violates, or provides inadequate steering to stay within the lane), the driver may put the vehicle in an unsafe situation.

The trial supervisor is responsible for maintaining safe operation of the trial and is expected to intervene in emergence situations. Consequently, an unsafe control action was identified (UCA-11) in which the trial supervisor does not provide a control action (does not intervene) and causes a hazard. Moreover, the trial supervisor can also provide a hazard and put the vehicle in an unsafe situation when s/he intervenes, which led to the definition of UCA-12.
Finally, UCA-14 and UCA-15 display two contexts in which automation provides control of the vehicle and creates hazards.

Table 26 - Examples of unsafe control actions for the second STPA analysis

Controller	Control action (CA)	Not providing the CA causes hazard	Providing the CA causes hazard
Driver	Provide control of the vehicle		**UCA-6:** The driver provides inadequate control of the vehicle during manual mode
Trial supervisor	Intervene	**UCA-11:** The trial supervisor does not intervene when safety is threatened	**UCA-12:** The trial supervisor intervenes and puts the vehicle in an unsafe situation
Automation	Provide control of the vehicle		**UCA-14:** Automation provides control of the vehicle during manual driving
			UCA-15: Automation provides inadequate control of the vehicle during automated driving

Safety requirements

The 18 unsafe control actions identified in the first step of the STPA analysis were translated into safety requirements as illustrated in the following examples:

- **UCA-6:** The driver provides inadequate control of the vehicle during manual mode.
- **SR-6:** The driver must provide adequate control of the vehicle during manual mode.

- **UCA-11:** The trial supervisor does not intervene when safety is threatened.
- **SR-11:** The trial supervisor must intervene when safety is threatened.

- **UCA-12:** The safety supervisor intervenes and puts the vehicle in an unsafe situation.
- **SR-12:** The safety supervisor must not put the vehicle in an unsafe situation when s/he intervenes.

- **UCA-14:** Automation provides control of the vehicle during manual driving.
- **SR-14:** Automation must not provide control of the vehicle during manual driving.

- **UCA-15:** Automation provides inadequate control of the vehicle during automated driving.
- **SR-15:** Automation must provide adequate control of the vehicle during automated driving.

Scenarios leading to unsafe control actions and refined safety requirements (STPA step 2)

The second step of the STPA analysis involves the elaboration of scenarios leading to the 18 identified unsafe control actions and the definition of refined safety requirements. The first STPA analysis on the vehicle trial process involved high-level controllers and consequently the entire set of control flaws classification was not always relevant; the identified potential causes for unsafe control actions were mainly due to inadequate process models and inadequate feedback. Conversely, the second STPA analysis presented in this section, examines lower-level controllers at the operational process and therefore multiple categories of the control flaw classification (displayed in figure 19) such as the inadequate operation of vehicle sensors and vehicle actuators, ineffective control actions, incorrect process models, etc. were relevant to the elaboration of scenarios. The analysis of the control flaws leading to the 18 identified unsafe control actions for the vehicle trial operation, resulted in the generation of 38 scenarios and the definition of 51 refined safety requirements.

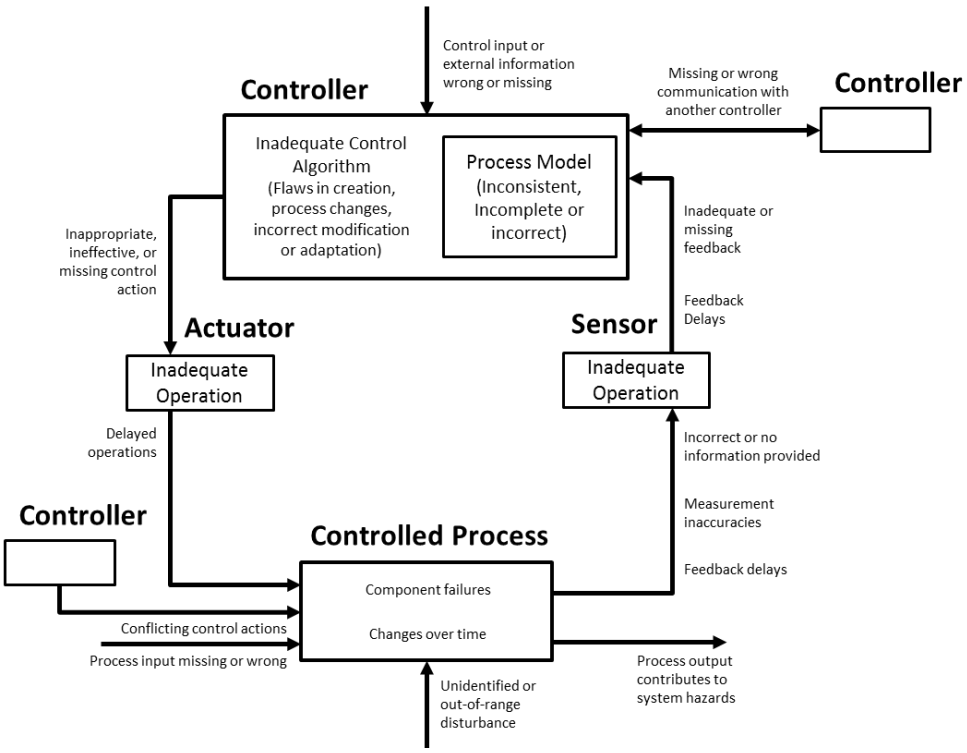


Figure 19 – Potential control flaws related to the control loop

Table 27 shows some examples of the scenarios and refined safety requirements generated for the UCA-15: Automation provides inadequate control of the vehicle during automated driving. (See appendix B for all the results of the STPA analysis on the vehicle trial operation involving a highway pilot system).

The first three scenarios include control flaws categories in the right side of figure 19 such as inadequate feedback, inadequate process model and inadequate control actions generated by the control algorithm which lead to the inadequate control of the vehicle. The three resulting refined safety requirements target the verification of automation's perception system and the feedback provided to automation by vehicle sensors, of automation's driving environment representation, and of the control actions and behavior that automation has on the trial route.

On the other hand, the last scenario includes the inadequate actuator operation control flaw category which is found in the left side of figure 19. In this scenario, an appropriate control action may have been generated by the control algorithm; however, the inadequate operation of the actuators that implement the control actions leads to inadequate control of the vehicle.

Subsequently, a refined safety requirement in which the trial supervisor overrides automation, was established.

Table 27 - Examples of scenarios and refined safety requirements defined for the second STPA analysis

UCA-15: Automation provides inadequate control of the vehicle during automated driving		
Control flaws	**Scenarios**	**Refined safety requirements**
Inadequate feedback: Driving environment	Automation provides inadequate control of the vehicle during automated driving because the feedback provided by vehicle sensors on the driving environment is inadequate	**SR-15.1:** The trial design team must verify that the perception system of the automated driving system provide adequate feedback on the driving environment
Inadequate model: Driving environment	Automation provides inadequate control of the vehicle during automated driving because automation has an inadequate representation of the driving environment	**SR-15.2:** The trial design team must verify that automation has an adequate representation of the driving environment
Inadequate control actions: control of the vehicle	Automation provides inadequate control of the vehicle during automated driving because the control algorithm generates inappropriate control actions on the vehicle actuators	**SR-15.3:** The trial design team must conduct pre-trials in which the vehicle is operated on AD mode on the trial route to validate that automation executes adequate actions to control the vehicle
Inadequate actuator operation: vehicle actuators	Automation provides inadequate control of the vehicle during automated driving because the vehicle actuators that implement control actions to control the vehicle have an inadequate operation	**SR-15.4:** The trial supervisor must intervene and override automation when automation provides inadequate control of the vehicle

The safety requirements and refined safety requirements defined through the two STPA analyses were the inputs for the framework to ensure the safety of automated driving trials. The process to structure these requirements into the five sections of the framework is described below.

4.4.3 Framework to ensure the safety of automated driving trials

The framework to ensure the safety of automated driving trials is the result of classifying the safety requirements defined through the STPA analysis into five sections. The results of the first STPA analysis on the vehicle trial process (17 safety requirements and 41 refined safety requirement) were organized to create sections 1-4; and the results of the second analysis on an automated driving trial involving a highway pilot system (the 18 safety requirements and 51 refined safety requirement) were structured to create section 5.

Framework sections 1-4 based on the safety requirements of the vehicle trial process

The following sections were defined according to the joint-responsibilities of the controllers involved in the multiple phases of the design and development of a vehicle trial.

1. Definition of policies and resources for the development of vehicle technology and vehicle trials: the government, funding agencies and company management define the policies and resources that influence the context for the design and development of vehicle technology and the vehicle trials. For instance, the government sets regulations on vehicle certification and open road vehicle testing. Furthermore, funding agencies such as the European Commission provide resources to conduct field operational trials. Lastly, the company management of vehicle manufacturers provides resources and company policies regarding vehicle trials.
2. Establishing orientations for the development of vehicle technology and vehicle trials: the company management sets the roadmap that guides the development of future vehicle technology and future vehicle trials. Additionally, funding agencies can also push vehicle manufacturers to develop and test specific technology. For example, the European Commission can make a call for projects on vehicle technologies addressing vulnerable road users like pedestrians and two wheelers.
3. Approval of vehicle trials: automakers or any organization conducting automated driving trials on open roads must request an authorization from the French government. Moreover, vehicle manufacturers have an in-house procedure for the approval of vehicle trials in which several controllers intervene like the department providing the prototype, experts, etc.
4. Design and development of vehicle trials: the last section comprises the controllers within the company who are directly responsible for the design and development of a vehicle trial. They prepare and organize the trial, identify trial hazards and safety measures and request the approval of the trial.

As observed in figure 33, sections 1-4 of the framework have a hierarchical relationship with control structure; the sections are related to the levels of the controllers and increase as the process gets closer to the vehicle trial operation.

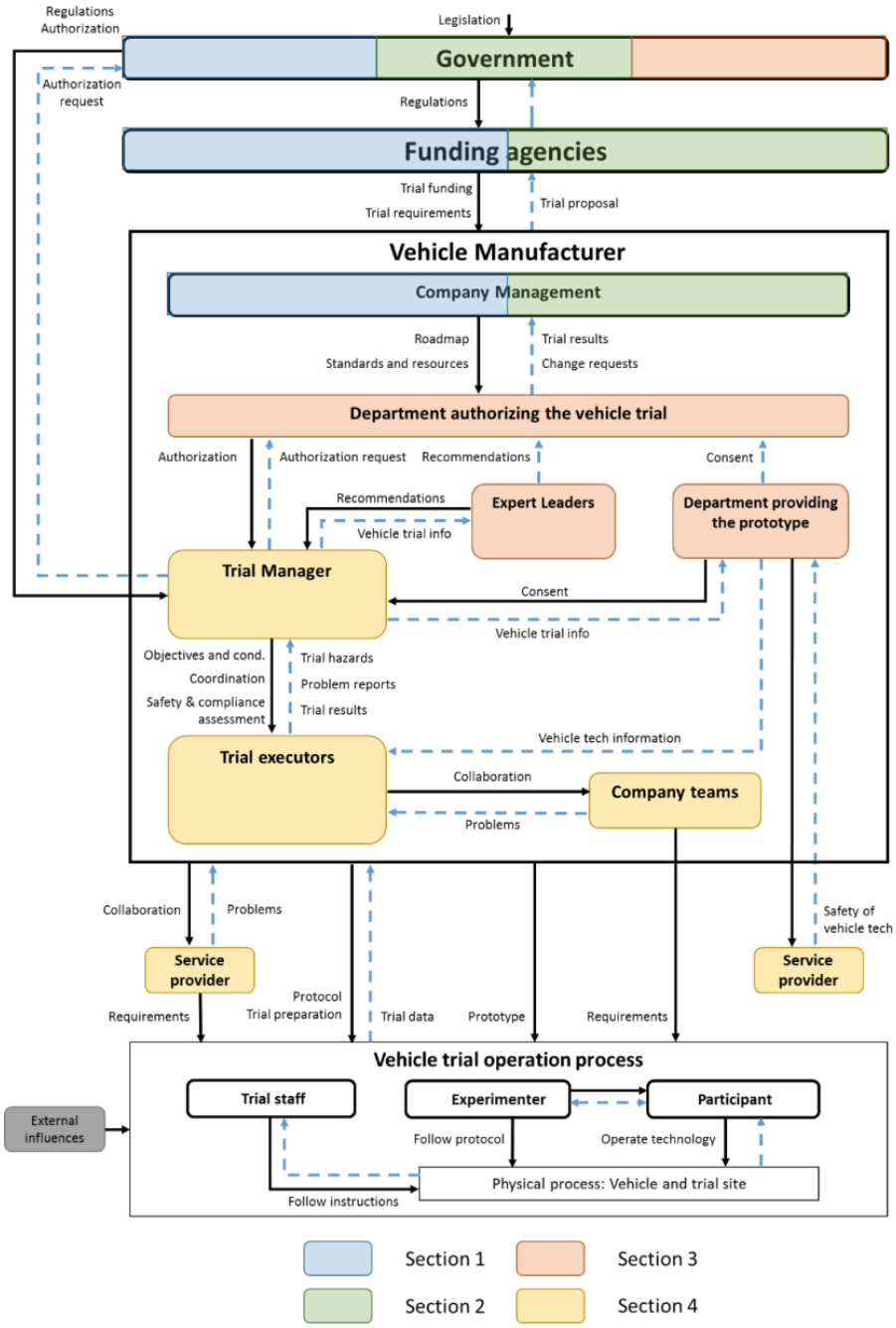


Figure 33 - Sections 1 to 4 of the framework relative to the control structure

The colors around the multiple controllers indicate the sections of the framework in which controllers are involved.

Next, clusters were generated to group the refined safety requirements and safety requirements of the controllers involved in each section. The clusters were used to create categories within the 4 sections.

Table 28 shows an example of the clusters and categories established for section 3 approval of vehicle trials, which concerns the government, the department authorizing the trial, company experts and the department providing the prototype.

Table 28 - Examples of clusters A to B and categories created for section 3

Section 3: Approval of vehicle trials			
Refined safety requirements	**Safety requirements**	**Clusters**	**Categories**
RSR-2.1: The government must have an adequate model of the vehicle trial	**SR-2:** The government must not authorize unsafe vehicle trials	A	Adequate model of trial safety and compliance
RSR-7.1: The department authorizing the vehicle trial must have an adequate representation of the compliance and safety of the trial	**SR-7:** The department authorizing the vehicle trial must not authorize a non-compliant or/and unsafe vehicle trial		
RSR-8.2: The company experts must have an adequate model of the vehicle technology and the vehicle trial	**SR-8:** The company experts must provide adequate recommendation to ensure the compliance and safety of the trial		
RSR-9.1: The department providing the prototype must have an adequate model of the prototype's level of maturity and safety	**SR-9:** The department providing the prototype must not give consent to use the prototype in an unsafe trial		
RSR-9.3: The department providing the prototype must have an adequate model of the vehicle trial			
RSR-2.2: The trial manager must provide adequate feedback in the dossier for a trial authorization request	**SR-2:** The government must not authorize unsafe vehicle trials	B	Adequate feedback on trial safety and compliance
RSR-7.2: The trial manager, company experts and department providing the prototype, must provide the department authorizing the vehicle trial with adequate feedback on the compliance and safety of the trial.	**SR-7:** The department authorizing the vehicle trial must not authorize a non-compliant or/and unsafe vehicle trial		
RSR-8.3: The trial manager must provide adequate feedback on the vehicle technology and vehicle trial in the recommendation request	**SR-8:** The company experts must provide adequate recommendation to ensure the compliance and safety of the trial		
RSR-9.2: The service providers that participate in the development of the trial must provide adequate feedback on the prototype's level of maturity and safety	**SR-9:** The department providing the prototype must not give consent to use the prototype in an unsafe trial		
RSR-9.4: The trial manager must provide adequate feedback on the vehicle trial to the department providing the prototype			
RSR-8.1: The company management must explicitly incorporate the function of providing recommendations for vehicle trials into the company expert's job functions	**SR-8:** The company experts must provide adequate recommendation to ensure the compliance and safety of the trial	C	Adequate commitment and knowledge for the approval process
RSR-8.4: The company experts must have an adequate model of the frameworks applicable to vehicle trials			

The approach illustrated in table 28 was applied to all the safety requirements and refined safety requirements identified for the vehicle trial process. An overview of the results of sections 1-4 of the framework is illustrated in Figure 34. It displays the controllers concerned by each section (illustrated in grey boxes), the categories identified based on the clusters (illustrated in white boxes) and the requirements and refined safety requirements issued from STPA analysis covered in each category (illustrated according to the color coding of figure 33).

Section 1: Definition of policies and resources for the development of vehicle technology and vehicle trials

This section covers the safety requirements that must be enforced by the government, funding agencies and company management, in order to provide adequate policies and resources for the development of vehicle technology and vehicle trials; policies and resources represent binding-guidelines and means to influence and regulate the development of safe vehicle technologies and safe vehicle trials. Based on the clusters, four categories of safety requirements were defined for section 1.

- Adequate model of the relevance of existing policies and resources for the vehicle trials being conducted: The controllers in this section need to be aware whether or not the current policies and resources support the safety of the vehicle trials; and whether or not they need to be modified or even completely changed.
- Adequate model of the vehicle technology being tested: In order to define policies and resources, controllers need to have a correct and consistent representation of the vehicle technology being tested, the differences relative to existing vehicle technology and the main hazards associated to the technology.
- Adequate model of the vehicle trials: controllers also need to be aware of the trial conditions and hazards associated to the vehicle trial.
- Adequate feedback on vehicle trials: lastly, the controllers in section 1 must receive adequate feedback from the lower-level controllers regarding vehicle trials. This feedback helps higher-level controllers to have adequate representations regarding the three previous categories.

Section 2: Establishing orientations for the development of vehicle technology and vehicle trials

This section groups the safety requirements that the funding agencies and the company management must enforce in to define orientations (roadmaps and trial conditions) that contribute to the development of safe vehicle technologies and safe vehicle trials. The safety requirements and refined safety requirements in section 2 were classified in four categories:

- Adequate model of orientation's safety: the funding agencies and the company management must have a consistent representation of the orientations' safety. They need to know if the directions set for vehicle technology and vehicle trials are unsafe.
- Adequate feedback on the hazards related to the established orientations: the lower levels of the system must provide funding agencies and company management with adequate feedback regarding the hazards of the vehicle technology and vehicle trial orientations.
- Values regarding safety and innovation: the company management must evaluate the value that they assign to safety relative to innovation. Are they willing to push the development of unsafe vehicle technology for the sake of innovation and competition with other automakers?
- Clear orientations and adequate orientations' dissemination: the company management needs to define a clear and understandable roadmap and distribute it to all the employees involved in the development and testing of vehicle technology.

Section 3: Approval of vehicle trials

The third section comprises the safety requirements that must be enforced by the government and several levels of the company organizing the vehicle trial, to approve or authorize safe and compliant vehicle trials. The safety requirements concerned by this section were classified into three categories:

- Adequate model of the compliance and safety of the vehicle trial: the government and the company controllers in charge of the approval of the trial need to have an adequate model of the safety and compliance of the trial in order to authorize it.
- Adequate feedback on the compliance and safety of the vehicle trial: the controllers of section 3, must receive adequate feedback on compliance and safety of the trial from lower-level controllers. This feedback helps the higher-level controllers in charge of trial approval to decide (or not) to authorize the trial.
- Adequate commitment and knowledge to support the approval process: the company controllers that contribute to trial approval must be committed to the process and

dispose of the necessary knowledge to evaluate whether or not the trial can be authorized.

Section 4: Design and development of vehicle trials

The fourth section covers the safety requirements enforced by the trial manager, trial executor(s), and the company teams and service providers participating, to design and develop safe and compliant vehicle trials. Furthermore, the section was divided into three sub-sections.

Sub-section 4.1: Trial organization and preparation

The trial manager, trial executors, company teams and service providers work together to organize and prepare the trial. Three categories were identified to group the safety requirements within this sub-section:

- Adequate project management.
- Adequate model of the requirements to prepare a trial.
- Adequate verification during trial preparation.

Sub-section 4.2: Trial data

As a part of the trial's design and development, the trial executor must establish the data that the trial will record. Two categories were defined for this subsection:

- Adequate model of data needs for the trial's objectives and liability matters.
- Adequate verification of the trial data.

Sub-section 4.3: Safety and compliance of the trial

Finally, the safety and compliance of the trial must be evaluated during the development of the trial. The safety requirements contained in this subsection were grouped into four categories:

- Adequate model of vehicle technology and vehicle trials.
- Adequate model of the requirements for trial safety and compliance.
- Adequate feedback regarding trial safety and compliance.
- Adequate specifications for trial operations.

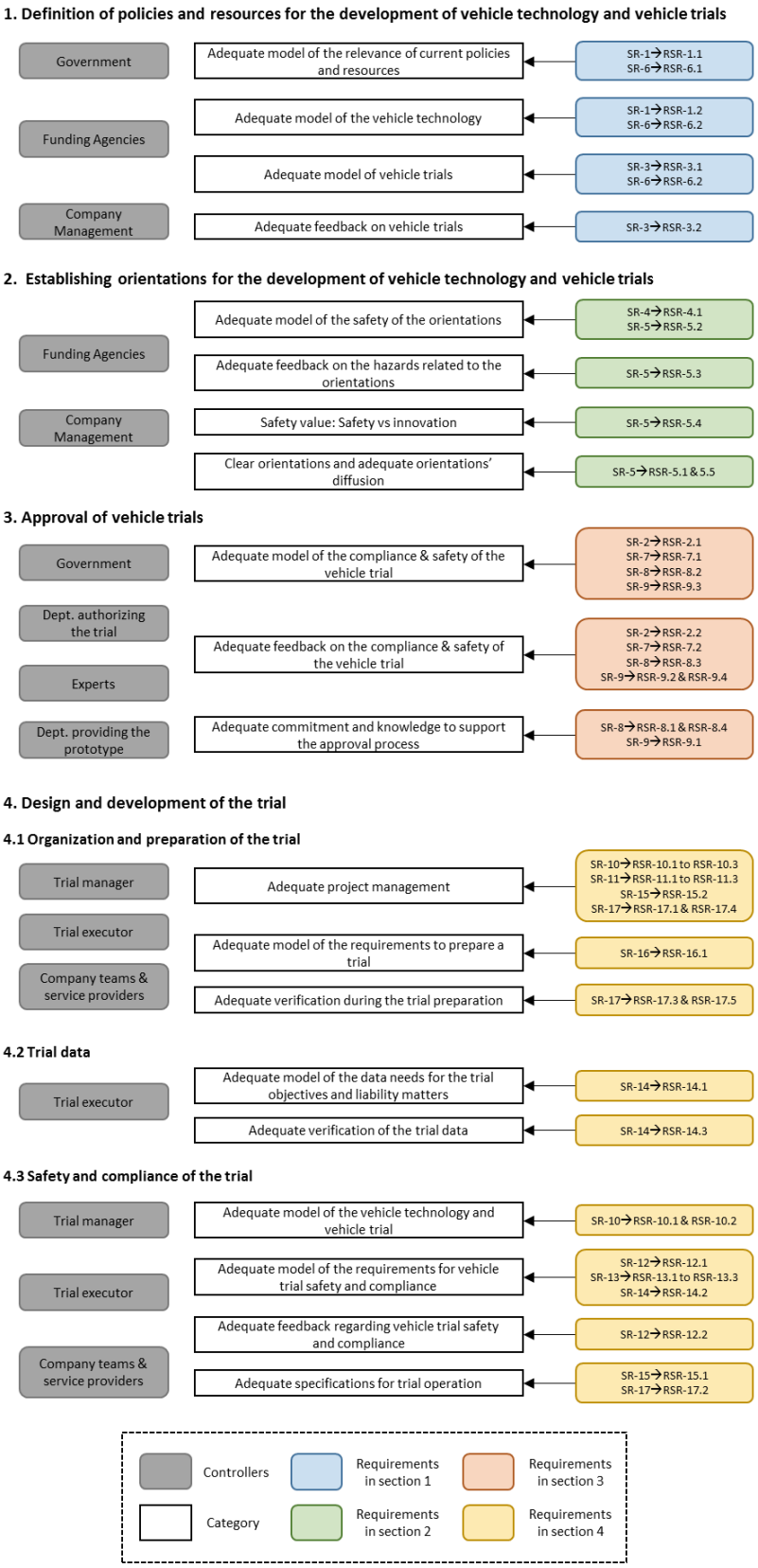


Figure 34 - Overview of the Framework sections 1 to 4

Framework section 5 based on the safety requirements on the automated driving trial operation

The last section of the framework organizes the safety requirements issued from the second STPA analysis on an automated driving trial operation involving the highway pilot system. As opposed to the vehicle trial process, which had several joint-responsibilities that needed to be identified to structure safety requirements, the control structure of the automated driving trial operation has one main responsibility: to ensure the safety of the vehicle trial operation. Consequently, there was no need to analyze the control structure in order to define several sections; instead, the only one section for the vehicle trial operation was established and directly divided into three sub-sections:

1. Safety related to the maturity of the vehicle technology being tested.
2. Safety related to the vehicle trial operation.
3. Trial data.

Following the approach of sections 1-4, clusters were generated to group the refined safety requirements and safety requirements of the three sub-sections. Subsequently, the clusters were used to create categories within the three sub-sections.

Table 29 shows an example of the clusters and categories established for section 5.2 (Safety related to the vehicle trial operation), which concerns trial staff, trial experimenter, driver participant, trial supervisor and automation.

The approach illustrated in table 29 was applied to all the safety requirements and refined safety requirements identified for the automated driving trial operation in order to structure the three sub-section of section 5.

Figure 35 displays an overview of the results of section 5; the controllers are illustrated in grey boxes, the categories identified based on the clusters are illustrated in white boxes and the and the requirements and refined safety requirements issued from STPA analysis covered in each category are illustrated in purple boxes.

Table 29 - Examples of clusters and categories created for section 5.2

Section 5.2: Safety related to the vehicle trial			
Refined safety requirements	**Safety requirements**	**Clusters**	**Categories**
RSR-1.1: The trial design team must provide the trial staff with adequate, correct, complete and understandable trial instructions.	**SR-1:** The trial staff must adequately follow instructions to manage logistics, to secure the trial and to ensure the safety of people involved in the trial	A	Adequate trial instructions and protocol
RSR-2.1: The trial design team must provide the trial experimenter with an adequate, correct, complete and understandable trial protocol.	**SR-2:** The trial experimenter must adequately follow the trial protocol		
RSR-3.1: The design team must provide the trial experimenter with adequate feedback on the driver participant instructions	**SR-3:** The trial experimenter must adequately provide instructions to the participant (how to operate the vehicle technology, what to do, safety instructions, etc.)		
RSR-5.1: The trial experimenter protocol must limit the interactions with the driver participant that generate stress and distraction	**SR-5:** The trial experimenter must not cause the driver participant to put the vehicle in an unsafe situation when the trial experimenter interacts with the driver participant		
RSR-6.2: The driver must have an adequate model of the current driving mode and the HMI interfaces for the two driving modes	**SR-6:** The driver must provide adequate control of the vehicle and comply with traffic rules during manual driving mode	C	Adequate model and training of trial supervisor and driver
RSR-7.2: The driver participant training must cover the AD engagement procedure, notably when to release control of the vehicle	**SR-7:** The driver releases the control of the vehicle before the automated driving system is engaged		
RSR-8.3: The driver participant training must cover the HMI information and sequences to transition to AD mode	**SR-8:** The driver must not release the control of the vehicle after the automated driving system engagement		
RSR-9.3: The driver participant training must cover the importance of regaining situation awareness before responding to the takeover request	**SR-9:** The driver must not put the vehicle in an unsafe situation when s/he has not regained situation awareness		
RSR-10.2: The driver participant training must cover how to respond to the takeover request	**SR-10:** The driver should respond to the takeover request		
RSR-11.1: The trial supervisor must perceive the traffic environment when s/he intervenes	**SR-11:** The trial supervisor must not put the vehicle in an unsafe situation when s/he intervenes		
RSR-11.2: The trial supervisor must have an adequate model to determine which situations need intervention and to adequately intervene (operate the override actuators, emergency switch and stop button)			
RSR-11.3: The trial supervisor must be trained to learn how to execute appropriate control actions when s/he intervenes			

Section 5: Vehicle trial operation

As seen in figure 35, section 5 includes the safety requirements enforced by the trial staff, trial experimenter, driver participant, trial supervisor, and automation to ensure the safety of the vehicle trial operation. Section 5 was further divided into three sub-sections: safety related to the maturity of the vehicle technology being tested, safety related to the vehicle trial operation and trial data.

Sub-section 5.1: Safety related to the maturity of the vehicle technology being tested

This sub-section structures the safety requirements related to the maturity level of the vehicle technology being tested into 6 categories:

- Adequate HMI operation and correct HMI information: the driver participant needs to receive adequate feedback from the highway pilot system's prototype regarding the status of the driving mode, ADS engagement, takeover requests, etc. in order to safely operate the highway pilot system.
- Adequate actuator operation: vehicle actuators need to have an adequate operation to enable the driver to safety operate the prototype and to allow the trial supervisor to intervene when safety is threatened.
- Adequate operation of automation's perception system: the perception system of the vehicle must have an adequate operation to detect when the ADS is available and when takeover requests are needed, and to provide control of the vehicle.
- Adequate automation's model of the driving environment and the operational design domain: automation must have an adequate representation of the driving environment and its operational design domain to provide control of the vehicle and detect the need for takeover requests.
- Adequate actions executed by automation: automation must execute control actions to provide adequate control of the vehicle, to send takeover requests to the driver and to provide minimal risk maneuvers for fallback performance.
- Verification of the maturity level of the vehicle technology via pre-tests: the maturity level of the prototype has to be verified before the trials.

Sub-section 5.2: Safety related to the vehicle trial operation

Sub-section 5.2 organizes the safety requirement issued from the STPA analysis that the trial staff, the trial experimenter, the driver participant, the trial supervisor and automation need to enforce to ensure the safety of the trial at an operational level. These requirements are classified into 4 categories:

- Adequate trial instructions and protocol: all the human controllers of the trial operation need adequate trial instructions and a safe protocol in order to carry out their responsibilities and contribute to the safety of the trial.
- Adequate model and training of the trial staff and trial experimenter: the trial staff and trial experimenter must be trained and given all the necessary information to follow the trial instructions and the protocol.
- Adequate model and training of the driver participant and the trial supervisor: the driver and the trial supervisor must be trained and given all the necessary information to operate the prototype and carry out their responsibilities.
- Adequate responses and recovery actions: all the human controllers need to be aware of the responses that they need to have in case of unsafe vehicle behavior or an emergency. Moreover, the trial supervisor must be capable of intervening and providing adequate recovery actions.

Sub-section 5.3: Trial operation data

The last sub-section structures the requirements on the trial data that the trial experimenter needs to enforce, into two categories:

- Adequate feedback on how to record data: the trial experimenter must receive adequate instructions on how to record trial data.
- Verification of the data recording process during trial operation: the trial experimenter has to verify that the data recorder is recording data before the start of the driving phase of every trial.

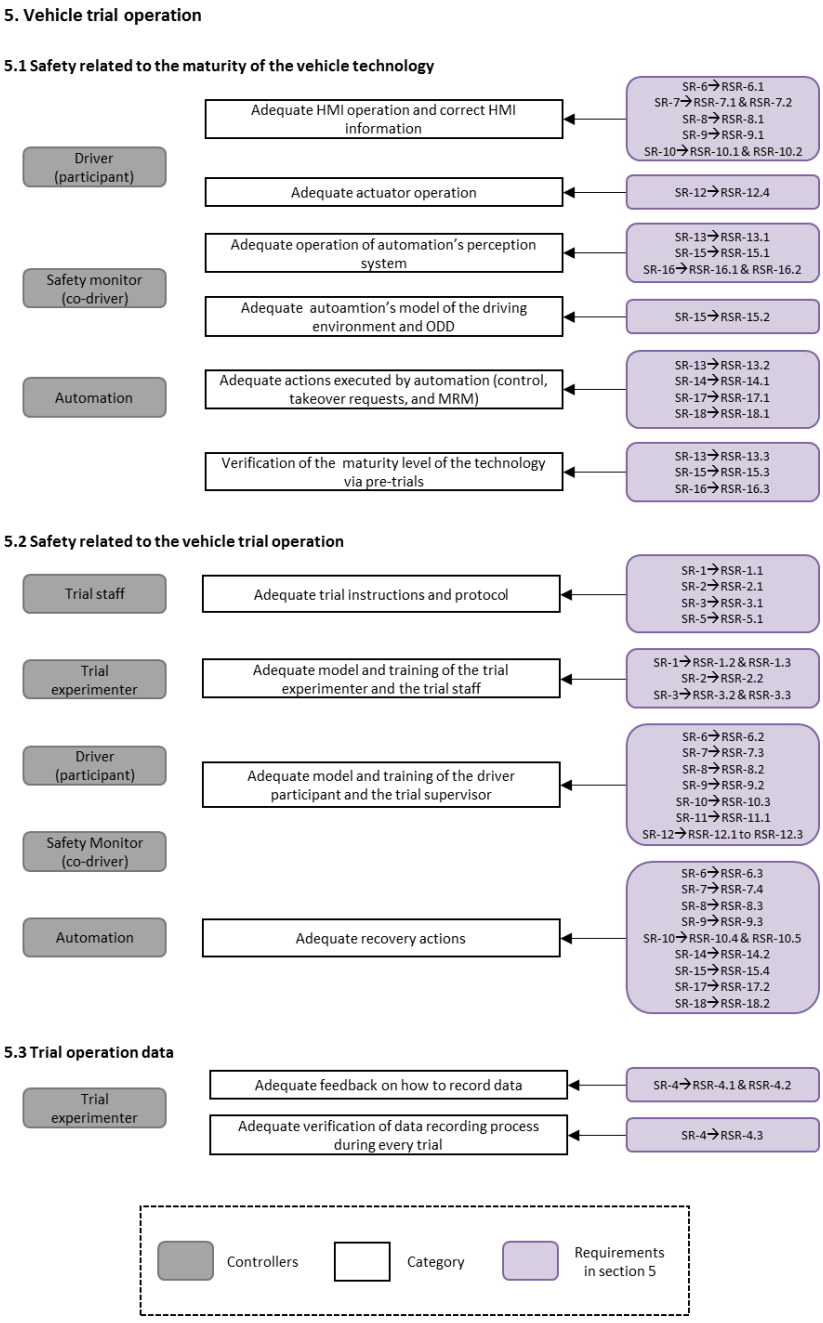


Figure 35 – Overview of the Framework section 5

4.5 Discussion

The aim of this chapter was to ensure the safety of automated driving trial processes by using the safety requirements resulting from two STPA analyses to establish a five section framework. The framework outlines the safety requirements that the controllers at the higher-levels of the vehicle trial system and the controllers at the operational level need to enforce in order to

conduct safe vehicle trials. The findings of this chapter are discussed relative to the scope and the contents of the framework.

4.5.1 The scope of the framework

The scope of framework developed in this chapter covers controllers across all the levels of the vehicle trial sociotechnical system, from the government and funding agencies, to the multiple actors within the company that design and develop vehicle systems (e.g. an automotive company), to the direct controllers of the physical system at the operational level. Conversely, other schemes for trial safety like the SAE guidelines (SAE International 2015), are mainly focused on the management of the trial drivers and on trial operation; the SAE guidelines do not consider the role of higher-level controllers within the company such as the company management, the departments that authorize the trial and the company experts. As a consequence, the SAE guidelines limit their scope to the responsibilities and safety measures on the management of trial drivers and on trial operation; they miss out on the opportunity to influence the safety of trial operations with safety measures enforced by higher-level controllers. For example, the SAE guidelines propose that the manager in charge of test drivers are responsible for explaining the organization's specific rules about test driving, however, the process to define those specific rules and the safety of the rules is not addressed. On the other hand, the framework introduced in this chapter contains safety requirements on providing feedback to the drivers regarding the test driving rules, and additional requirements on higher-level controllers involved in the definition of those rules, such as:

- **SR-6:** The company management must provide adequate standards and resources for vehicle trials.
    - **RSR-6.2:** The company management must have an adequate model of the vehicle technologies and vehicle trials.

The previous example illustrates the importance of extending the scope of the controllers that participate in trial safety beyond the operational process; exclusively focusing on the controllers at the sharp end of the process overlooks the influence of the controllers in the other parts of the process. Accordingly, it is recommended that the entire system be considered to identify the control mechanisms that enforce trial safety at all the levels of the system.

4.5.2 The contents of the framework

In terms of contents, the framework includes a larger set of subjects than the SAE guidelines. While the SAE guidelines primarily target the test driving training, the management of test drivers, the aspects to be considered in the trial risk management process, software considerations, the selection of trial routes and the trial data; the framework targets subjects like the definition of policies and resources that shape vehicle trials, the process of trial approval, the design and development of the trial which includes trial management and assessing the safety and compliance of the trial, the safety related to the trial and the maturity level of the prototype (containing software and hardware considerations), the training of the trial staff, experimenters and drivers and trial data. Although some of the additional contents of the framework such as the orientations of vehicle technology and vehicle trials can be explained by the larger scope, the framework contents regarding the lower levels of the system still addressed more subjects than the SAE guidelines. For example, the SAE guidelines did not mention the trial protocol design or the training of the trial staff.

Limitations

Like in all analyses, the results of an STPA analysis depend on the model i.e. representation of the system being examined. Therefore, a limitation of the framework is the level of vehicle trial experience and system understanding of the person(s) performing the STPA analysis. Whilst assistance was provided during the analysis (e.g. control structures were validated by experts on the system) the active participation of experts through the entire STPA analysis may enhance the process and allow the identification of additional safety requirements.

A second limitation is the specificity of the analysis which is related to the system being considered. In fact, the trial processes conducted in other countries and by other vehicle manufacturers can differ from the vehicle trial in France involving Renault. For example, the regulations established by the government on vehicle trials or the trial approval procedure within a company may be different. Furthermore, different operational processes for instance a vehicle trial with another automated driving systems and expert drivers in adverse conditions

may involve different safety requirements. Consequently, the applicability of the framework to other sociotechnical systems and other trial operations needs to be examined.

4.6 Conclusion

This chapter examined how STPA contributes to ensure the safety of the entire automated driving trial process by providing a five section framework based on the safety requirements identified through two STPA analyses: a first analysis on the vehicle trial process and a second analysis on an automated driving trail operation involving a highway pilot system.

The framework illustrates the multiple interactions, responsibilities and requirements of the controllers at all the levels of vehicle trial system. When compared with the SAE guidelines for the safety of automated driving trials, the framework developed in this chapter displays a larger scope and more comprehensive contents.

4.6.1 Future work

Three future research applications were identified:

1. The whole STPA analysis should be conducted with the support and active participation of several experts on the vehicle trial process and vehicle technology being tested. The results of the STPA performed with the experts could validate the methodology and findings of this chapter at the conceptual level. Moreover, the vehicle trial considered in the STPA analysis will take place in the second semester of 2017, providing the opportunity to further validate findings at an empirical level.
2. The STPA analysis at the vehicle trial process should be conducted on other vehicle trial systems to investigate how other control structures and corresponding safety requirements can differ from the one examined in this chapter. Additionally, the STPA analysis at the trial operation level involving other automated driving systems and trial conditions should also be explored in order to further verify the applicability of the methodology and the approach. To this end, the framework will be applied to the upcoming automated driving trials conducted by Renault, the vehicle trials of the L3Pilot project.

Résumé chapitre 5: Analyse des accidents de la route impliquant la conduite automatisée

---

Le chapitre 5 aborde la troisième question « Comment analyser les accidents de la route impliquant des systèmes de conduite automatisée ? » en élaborant une nouvelle méthode d'analyse des accidents impliquant un ou plusieurs véhicules autonomes nommée CASCAD. Tout d'abord, des éléments spécifiques ont été identifiés dans les méthodes existantes dans la sécurité routière, HFF et DREAM. Ensuite, des éléments explicatifs ont été développés en utilisant des concepts du modèle STAMP afin de faciliter l'application de CAST sur les accidents de la route impliquant la conduite automatisée. Tous ces éléments ont été intégrés à CAST pour créer la nouvelle méthode CASCAD.

CACSCAD a enfin été appliquée à un cas réel d'accident impliquant un véhicule autonome, celui de la Tesla datant de mai 2016.

Chapter 5: CASCAD—an accident analysis method for crashes involving automated driving

5.1 Chapter overview

Chapter 5 introduces an accident analysis method for crashes involving automated driving called CASCAD. As observed in figure 36, this chapter identifies two road safety specific elements from existing road crash analysis methods (i.e. HFF and DREAM), and develops three guidance elements using concepts from STAMP and a STAMP-based analysis; these elements are incorporated into CAST to create CASCAD. Next, the CASCAD method is illustrated with the available data on a Tesla crash that occurred on May 2016. Lastly, the findings of the chapter are discussed.

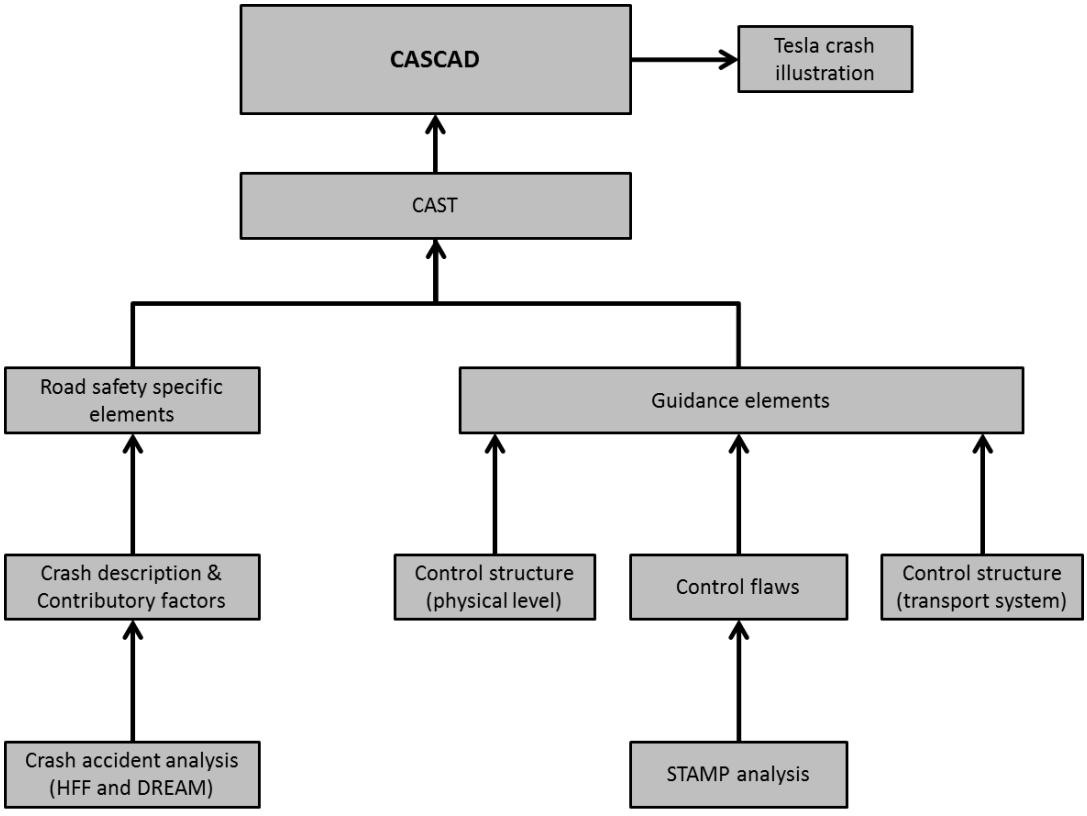


Figure 36 – CASCAD

5.2 Introduction

Although automated driving is expected to reduce the number of crashes by removing the human driver error, the changes brought by the whole range of vehicle automation levels into the road transport system may not eliminate existing causal factors and it could even introduce new causal factors. For instance at the driver assistance level (SAE 1), the driver has to execute either lateral or longitudinal control and therefore the causal factors related to the traditional driving task, such as alcohol consumption, distraction, fatigue, inexperience, etc. are still present. Moreover, the levels in which the driver delegates the monitoring of the driving environment to automation and is requested to intervene when automation can no longer assure safe operation (SAE 3 and SAE 4), may introduce causal factors associated to the loss of driver's situation awareness (N. Merat and Jamson 2009; Natasha Merat et al. 2014). Furthermore, all the levels of vehicle automation can introduce causal factors related to the driver's skill degradation and overreliance on automation and to the vehicle sensors' operation (Natasha Merat and de Waard 2014; NHTSA 2014; Cunningham and Regan 2015; Kyriakidis et al. 2017).

In terms of interactions among the road users, the mixed traffic conditions in which automated vehicles and non-automated vehicles share the same traffic environment, can introduce causal factors related to the limited capability of automated vehicles to comply with traffic rules and to behave like human drivers. Additionally, automated vehicles may also bring causal factors associated to the limited capability of automation to communicate via gestures (Maurer et al. 2016). Consequently, the road safety community has to prepare for the analysis of crashes involving automated driving in order to monitor the state of road safety and to identify effective prevention strategies for future road transport systems which include automated vehicles. To this end, the road safety community must find appropriate accident analysis methods which can assist the understanding of the entire set of factors related to crashes involving automated driving.

On the other hand, several researchers have suggested that the traditional methods are no longer enough to significantly improve road safety and that a change of paradigm towards systems theory is needed (Larsson, Dekker, and Tingvall 2010; Salmon and Lenné 2015; Hughes,

Anund, and Falkmer 2015; Hughes et al. 2015). Among the models which have a systems theoretical foundation, STAMP is an accident model that integrates concepts from systems theory to model and examine the interactions between the stakeholders at all the levels of the whole sociotechnical system (Leveson 2004, 2011). Moreover, STAMP was intendedly developed to deal with systems in which there is a high degree of automation, offering a framework that is suitable for the analysis of automation, human operators and organizations. Further, CAST (Causal Analysis based on STAMP) which is an accident analysis method developed based on STAMP, could constitute a suitable candidate for the analysis of crashes involving automated driving.

However, CAST is a generic method which is not specific to any industry; therefore it does not take into account the particularities of the road safety domain. As shown by Underwood the lack of industry-specific elements may prevent practitioners from adapting CAST (Underwood and Waterson 2013). (Underwood 2013) found that practitioners do not adopt systems accident analysis methods such as CAST and FRAM, because these methods do not meet their needs in terms of usability, graphical outputs and industry-specific taxonomies; he recommends providing more usage guidance material for a given industry. Accordingly, CAST should be adapted to include additional usage guidance regarding road safety specific-elements and automated driving.

5.2.1 Study aim and objectives:

The aim of this study was to tackle the third research question *"how to analyze road crashes involving automated driving?"* by extending CAST into a method called CASCAD (**C**ausal **A**nalysis using **S**TAMP for **C**onnected and **A**utomated **D**riving) which incorporates road safety-specific elements and automated driving, to assist a more complete analysis of crashes involving automated driving.

The following objectives were defined to achieve the aim of the study:

- Examine traditional accident analysis techniques from the road safety domain to identify road safety-specific elements that can be transferred to the analysis of crashes involving automated driving systems.

- Analyze an automated driving system at the operational level using STAMP to develop elements that facilitate the application of CAST on the analysis of crashes involving automated driving systems.
- Build CASCAD by incorporating the road safety-specific elements and the automated driving elements into CAST, and illustrate[14] its application with a real-world crash involving automated driving systems.

5.3 Methods

This section presents the methods used to identify the elements specific to road safety from existing crash analysis methods. Next, it presents the methods to develop guidance elements for the use of CAST on automated driving and road crash analyses. Lastly, it describes how these two types of elements were incorporated into CAST to create CASCAD, and how a real crash involving a Tesla was used to illustrate the application of CASCAD.

5.3.1 Elements specific to road safety

Two traditional crash analysis methods called the Human Functional Failure (HFF) framework (Van Elslande and Alberton 1997; Van Elslande 2000b) and the Driving Reliability Analysis Method (DREAM) (Sagberg and Transportøkonomisk institutt (Norway) 2008; Ljung Aust et al. 2012) were described according to four characteristics: their aim, their conceptual and empirical basis, their key notions and their process. Subsequently, two road-safety specific elements relevant to the analysis of crashes involving automated driving were identified. Lastly, the integration of such elements into CASCAD was discussed.

5.3.2 Elements to facilitate the application of CAST on automated driving

STAMP concepts were used to examine a generic automated driving system in order to develop three elements that facilitate the application of CAST on crashes involving automated driving. Firstly, a generic control structure of the interactions at the physical level was built in which the

[14] This study illustrates the application of CASCAD using limited data from a real crash involving an SAE 2 system (Tesla's Autopilot) which happened in May 2016, for pedagogic purposes ; it does not intend to provide an accident analysis for the crash.

interactions between vehicles and the infrastructure were characterized in terms of control actions, feedback and external influences. Secondly, the direct controllers (i.e. the human driver and automation) were added to the control structure and analyzed using the control flaws suggested by (Leveson and Thomas 2013) to establish control flaws associated to the human driver controller and the automated controller. Thirdly, a control structure of the entire sociotechnical system which includes indirect controllers was built; it incorporates high-level controllers, such as international stakeholders, the government, automotive industries, road infrastructure companies and driving schools.

5.3.3 CASCAD

The road safety-specific elements and the elements developed to facilitate the application of CAST on automated driving were incorporated into the CAST method to create CASCAD. Furthermore, the available data regarding the Tesla crash in May 2016 were collected from NHTSA official reports (National Transportation Board 2016; Habib 2017), Tesla website (The Tesla Team 2016), and unofficial articles (Lambert 2016; Singhvi and Russell 2016) to establish a limited description of the crash and illustrate the first four steps of CASCAD.

5.4 Findings

Section 5.4 presents the identified road safety specific elements, the developed elements to facilitate CAST application on automated driving and road crashes, and the CASCAD method. Additionally, it analyzes the available data on the Tesla crash in May 2016 using CASCAD as a means to illustrate the application of the method.

5.4.1 Road safety-specific elements

This sub-section describes the Human Functional Failure (HFF) Framework and the Driving Reliability Error Analysis Method (DREAM) which are the main accident analysis methods used by Renault to examine road crashes. Subsequently, the identified elements from the two methods which can be transferred to the analysis of crashes involving automated driving are explained.

5.4.1.1 Human Functional Failure (HFF) Framework

The Human Functional Failure (HFF) Framework is described according to four characteristics: aim of the method, conceptual and empirical basis, main elements and the analysis process.

Aims of the method

The Human Functional Failure (HFF) framework has two main aims; firstly, to better understand the processes and mechanisms involved in road users' human functional failures in order to contribute to more effective countermeasures. Secondly, to identify new research areas for road safety (Van Elslande and Alberton 1997; Van Elslande 2000a).

Conceptual and empirical basis

Van Elslande suggested the need to move away from seeing the errors of road users as the main cause of crashes towards considering human functional failures as the result of malfunctions among the components of the driving system (i.e. the road user, the environment and the vehicle) and their interactions. To create an accident analysis method that supports such view of road user errors, Van Elslande reviewed concepts from the literature such as Rasmussen's functional hierarchy (Jens Rasmussen 1986), Reason's view on human error (Reason 1990) and human information processing models in order to generate a general level "theoretical" classification of human functional failures categories.

Next, empirical data from in-depth road crash analyses were examined to refine the classification by defining a specific level of human functional failures types relevant to the driving context. Additionally, the classification was applied to a larger set of empirical data to characterize road crashes in terms of several parameters (e.g. human functional failure type, pre-accident situation, explanatory elements) and to aggregate them into generic scenarios for every human functional failure type.

Key elements

a. Sequential description of road crashes:

As seen in figure 37, crashes are described according to four sequential phases:

1. The driving phase which corresponds to the normal driving situation.

2. The rupture phase in which an unexpected event interrupts the normal driving situation an upsets the balance of the system.
3. The emergency phase in which the road user tries to return to a normal situation by engaging in an emergency maneuver.
4. The impact phase which comprises the crash and its consequences.

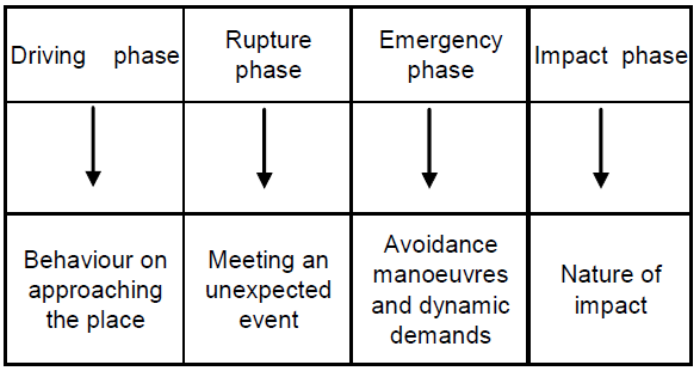


Figure 37 - Main phases of a road crash (Van Elslande and Fouquet 2007)

b. The HFF classification

The HFF classification (displayed in figure 38) includes two levels of human functional failures observed in road crashes. As seen in the left side of the figure in pale orange, the global level of the HFF classification includes six types of general failures categories associated to human failures in information acquisition, diagnosis, prediction, decision-making, action execution and overall failures. These six categories are inspired by the human information processing model, Rasmussen's functional hierarchy (Jens Rasmussen 1986) and Reason's distinction on errors and violations (Reason 1990). Furthermore, the specific level contains twenty types of failures associated to the driving activity (illustrated on the left side of the figure in pale blue) which are derived from the general failures and in-depth crash analysis data.

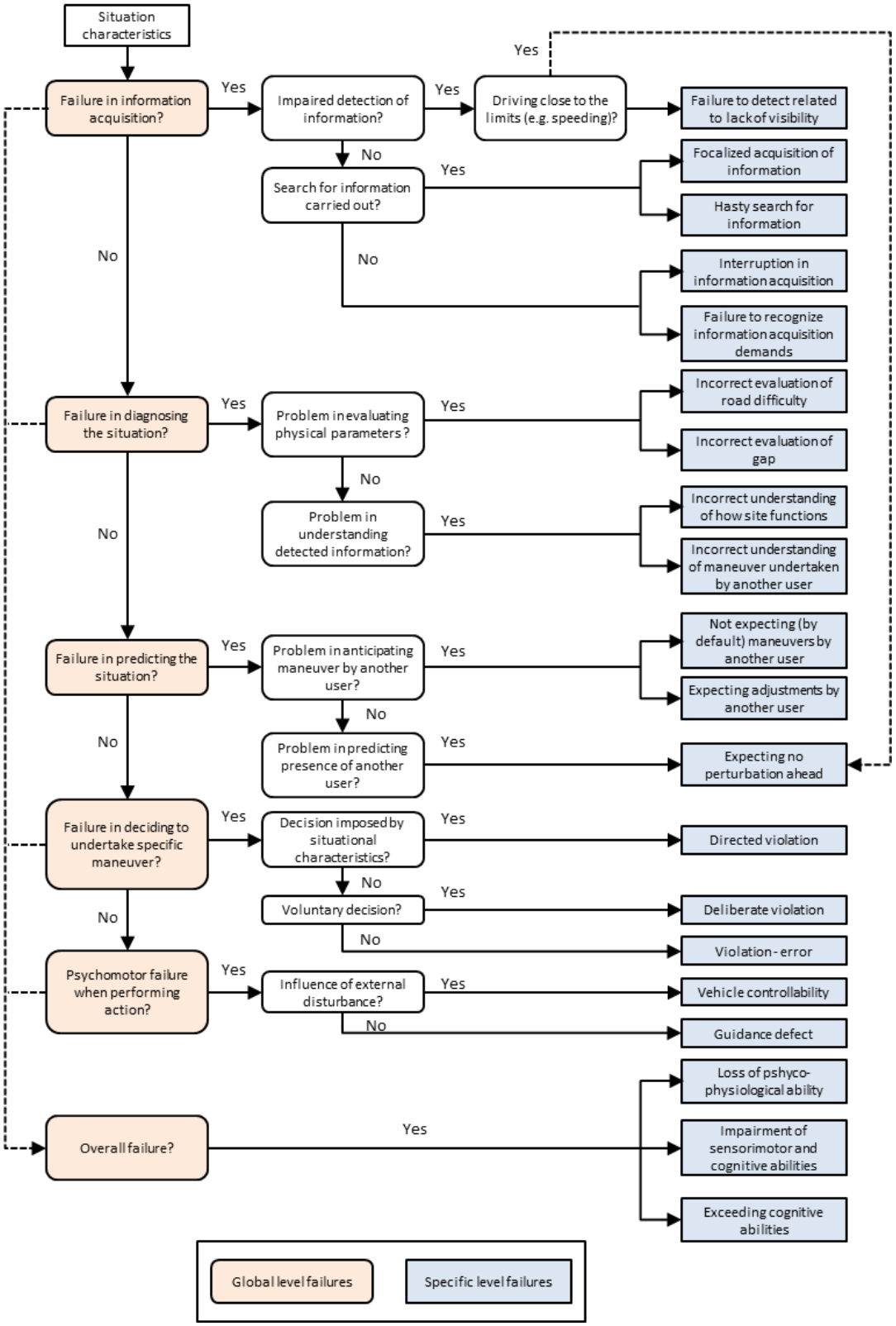


Figure 38 – Classification of the Human Functional Failures (Van Elslande and Fouquet 2007)

c. Explanatory elements:

The explanatory elements are the factors that contribute to functional failures. Van Elslande provided a list of explanatory elements classified into two main categories: endogenous to the driver and exogenous to the driver. Additionally, the endogenous category is divided into three sub-categories associated to the driver's state, the driver's experience and the task execution conditions, and the exogenous category is divided into three sub-categories related to the road, the traffic and the vehicle; table 30 illustrates examples of the explanatory elements sub-categories.

Table 30 - Examples of endogenous and exogenous element

Endogenous			Exogenous		
Driver state	**Experience**	**Task execution conditions**	**Road**	**Traffic**	**Vehicle**
Drowsiness Impairments Alcohol Alcohol/drugs Fatigue Distraction Stress	Over-experience Under-experience Episodic driving Unknown place Unknown vehicle Known route Known maneuver	Priority feeling Excessive trust Time constraints Secondary tasks Over-speeding	Traffic signs Limited visibility Site complexity Degraded road markings Luminosity	Difficulty to get a gap Pressure Punctual visibility problem Wind Other road user behavior	Tire pressure Degraded tire Suspension Mechanical failure Car's lighting

d. Generic scenarios:

The results of crash accident analyses using the HFF framework on 392 accident situations were aggregated to elaborate generic scenarios for the failures specific to the driving activity. The generic scenarios group the pre-accident situations at the rupture phase and the explanatory elements identified with the HFF framework. These scenarios have been used to identify the most frequent causal factors in a sample of crashes, to make-up for missing data, and to validate new crash analyses by comparing the results with the generic scenarios.

e. The degree of road user's involvement:

The degree of road user's involvement is determined by considering four categories:

1. Primary active in which the road user causes the disturbance.
2. Secondary active in which the road user is not the source of the disturbance but he/she does not try to resolve the conflict situation.

3. Non-active in which the road user is confronted to an atypical maneuver of others which is hardly predictable. However, if the non-active road user had anticipated the maneuver, s/he could have had avoided the crash.
4. Passive in which the driver is not involved in the destabilization of the situation; the passive road user cannot do anything to avoid the crash. For example when a vehicle gets rear-ended at a red light.

Analysis process

The HFF method combines its key elements to conduct a two-stage analysis of a road crash for each road user involved. The first stage consists of describing the crash as a sequence for each road user. The second stage comprises understanding the human functional failure(s) observed in the rupture phase of the crash by identifying the functional failures involved in the rupture phase of the crash, identifying the associated explanatory elements and determining the degree of involvement for each road user.

5.4.1.2 Driving Reliability Error Analysis Method (DREAM)

The second accident analysis method considered is the Driver Reliability Error Method (DREAM) which is the adaptation of the Cognitive Reliability Error Method (CREAM) to the driving activity. Like in the HFF framework, DREAM is described according to the aims of the method, its conceptual and empirical basis, its main elements and steps provided by the method to conduct an accident analysis.

Aim of the technique

The aim of the Driving Reliability Error Analysis Method (DREAM) is to gain a better understanding of failures between road users and contextual variables associated with the different parts of the system in order to contribute to accident prevention (Sagberg and Transportøkonomisk institutt (Norway) 2008).

Conceptual and empirical data

As in the HFF framework, the developers of the first DREAM version were interested in models and techniques that supported the shift of human error from the cause of accidents to the

consequence of the context in which the accident takes place. Accordingly, Hollnagel's Cognitive Reliability and Error Analysis Method (CREAM) (Hollnagel 1998) was selected as a basis of DREAM. The conceptual foundation of CREAM is underpinned by cognitive systems engineering, and more specifically by an alternative model of human information processing called Contextual Control Model (COCOM) that describes how competences (observation, interpretation, planning and execution) and performance depend on the context. Moreover, CREAM provides a classification scheme that links human error modes (called phenotypes) and contributory factors (called genotypes) associated to the human, technology, and organizations.

To adapt CREAM to the driving context and thus create the first version of DREAM, the classification scheme provided in CREAM was slightly modified to better fit the driving context; for instance, genotypes such as driving under the influence of alcohol, vehicle design and inadequate road design were incorporated into the classification scheme (Sagberg and Transportøkonomisk institutt (Norway) 2008). Subsequently, CREAM (with the slightly adjusted classification) was applied to analyze the data of fifteen in-depth crash analyses. Furthermore, posterior versions of DREAM have been developed to update the classification scheme by incorporating links between phenotypes supported by empirical evidence or research studies and to improve the methodology (Ljung Aust et al. 2012).

Key elements

a. Accident phases

As in the HFF, DREAM divides accidents into four phases: the driving phase, the discontinuity phase (rupture phase in the HFF method), the emergency phase and the crash phase.

b. Phenotypes

Phenotypes or manifestations constitute the first element of the classification scheme; they are the most relevant observable effects of human dysfunctional adaptive behavior in the discontinuity phase of a crash. They help investigators determine the moment when the human road user lost control in physical terms and the starting point of the accident analysis. As displayed in table 31, there are six categories of general phenotypes that describe dysfunctional adaptive behavior: timing, speed, distance, direction, force and object. Moreover, the general

categories are furthered divided into specific phenotype categories, for instance, timing can be divided into: too early action, too late action and no action.

Table 31 - List of phenotypes

Phenotypes	
General	Specific phenotype
Timing	Too early action ; too late action; no action
Speed	Too high speed ; too low speed
Distance	Too short distance
Direction	Wrong direction
Force	Surplus force; insufficient force
Object	Adjacent object

Some of the phenotypes are closely related, for example, if a vehicle collides with an oncoming vehicle during an overtaking maneuver, three phenotypes could explain the human dysfunctional behavior:

1. Timing, the driver may initiate the overtaking maneuver too early or too late.
2. Speed, the speed was too low to complete the overtaking maneuver.
3. Distance, the stretch of free road was too short to complete the overtaking maneuver. In this case, the crash investigator must consider all the collected data and select the most appropriate phenotype e.g. the specific phenotype is too late or too early action is more appropriate for a crash if the driver testimony states that s/he had not seen the other vehicle.

c. Genotypes

Genotypes are the second element of the classification scheme. They are the contributing factors that bring about the phenotypes (i.e. the causes of the observable effects). Usually, the genotypes cannot be observed and consequently they must be inferred or deduced from the collected data. They are classified according to Hollnagel's Man-Technology-Organization (MTO) model into driver-vehicle-traffic environment and organizational. As seen in table 32, the genotypes associated to the driver category are organized relative to problems with cognitive functions such as observation, interpretation, and planning in accordance to the COCOM model, as well as more general states of temporary and permanent person related factors including fatigue and permanent visual impairments (Hollnagel 1998). The genotypes of the

technology category include factors associated to the vehicle , such as problems with the Human Machine Interface and vehicle equipment failure and the traffic environment, and factors associated to the traffic environment. Lastly, the genotypes in the organization category include factors linked to the organization, maintenance, and vehicle and road design.

Table 32 - Overview of genotype categories

<table>
<tr><th colspan="3">GENOTYPE CATEGORIES</th></tr>
<tr><th colspan="2">Human Driver (B-F)</th><th>Technology (G-M)</th><th>Organization (N-Q)</th></tr>
<tr><td>Observation (B)
Interpretation (C)
Planning (D)</td><td>In accordance with COCOM</td><td rowspan="2">Vehicle (G-I)
Temporary HMI problems (G)
Permanent HMI problems (H)
Vehicle equipment failure (I)
Traffic environment (J-M)
Weather conditions (J)
Obstruction of view due to object (K)
State of road (L)
Communication (M)</td><td rowspan="2">Organization (N)
Maintenance (O)
Vehicle design (P)
Road design (Q)</td></tr>
<tr><td colspan="2">Temporary Personal Factors (E)
Permanent Personal Factors (F)</td></tr>
</table>

d. Links

The last element of the classification scheme is the links between the phenotypes and genotypes as well as between different genotypes. Links embody existing knowledge about how the phenotypes and genotypes can interact and be associated. The phenotype tables include the links between phenotypes and general genotypes, and the genotype tables include the links among genotypes.

For instance, table 33 illustrates an excerpt from the phenotype table in which the specific phenotype “Too early action (A.1.1)” in the consequents side of the table can be linked to general genotypes on the antecedents’ side of the table.

Table 33 - Excerpt from phenotypes table (Ljung Aust et al. 2012)

<table>
<tr><th colspan="4">PHENOTYPES</th></tr>
<tr><th>Antecedents (Causes)</th><th colspan="3">Consequents (Effects)</th></tr>
<tr><th>General Genotypes</th><th>Definition of GENERAL Phenotypes</th><th>Definition of SPECIFIC Phenotypes</th><th>Example for SPECIFIC Phenotype</th></tr>
<tr><td>Misjudgment of time gaps (C1)
Misjudgment of situation (C2)
Incomplete judgment of situation (C3)
Fear (E1)
Fatigue (E3)
Under the influence of substance (E4)
Sudden functional impairment (E6)
Temporary access limitation (G4)
Equipment failure (I1)
Strong side wing (J2)
Missed Observation (B1)
Late observation (B2)</td><td>Timing (A1)</td><td>Too early action (A.1.1)
The action is initiated too early, before the signal is given or the required conditions are established.</td><td>Intersection accidents
Starting from a standstill the driver passes the traffic light too early (before it tuned green).

Starting from a standstill the driver passes the stop/give away signal too early (before it the intersection is free).

The driver leaves his own lane too early – before the lane he is changing into is free.</td></tr>
</table>

For a driver that engaged an overtaking maneuver too early because the driver thought s/he had the time to overtake because s/he was tired, the general genotypes "Misjudgment of time gaps (C1)" and "Fatigue (E3)" can be linked to the "too early action" specific phenotype.

Next, the "INTERPRETATION (C)" genotype table displayed on table 34 can be used to link the "Misjudgment of time gaps (C1)" general genotype on the consequents side of the table to other general genotypes on the antecedents' side of the table. For instance, there can be a link between "Misjudgment of time gaps (C1)" and "Late observation (B2)".

Table 34 - Excerpt from DREAM's Interpretation genotype table (Ljung Aust et al. 2012)

INTERPRETATION (C) Interpretation includes for all but novice drivers, quick and automated (routine) procedures where typical situations and their associated actions are recognized and acted upon (script choice). Mistakes in interpretation occur at the sharp end – within the local event horizon			
ANTECEDENTS			**CONSEQUENTS**
GENERAL Genotypes	**SPECIFIC Genotypes (with definitions)**	**Examples for SPECIFIC Genotypes**	**GENERAL Genotypes (with definitions)**
Late observation (B2) False observation (B3) Attention allocation towards other than critical event (E2) Fatigue (E3) Under the influence of substances (E4) Psychological stress (E7) Permanent functional impairment (F1) Expectance of certain behaviors (F2) Expectance of stable road environment (F3) Habitually stretching rules and recommendations (F4) Overestimation of skills (F5) Insufficient skills/knowledge (F6) Incorrect ITS-information (G5) Reduced visibility (J1) Insufficient guidance (L1) Reduced friction (L2) Inadequate road geometry (L5) Inadequate transmission from road environment (M2) Unpredictable system characteristics (P4)	**Misjudgment of time gap due to incorrect** The driver misjudges the time gap due to a misjudgment of the approaching vehicle's speed	*Intersection* The driver is waiting to cross a street and assumes that the approaching car is keeping the 50 km/h speed limit. The car is, however, approaching at 70 km/h and as a result the driver overestimated the time gap he has to the approaching car. *Overtaking* The driver is overtaking another car when he suddenly realizes that he has underestimated the meeting's car speed and therefore also overestimated the available gap for the overtaking	Misjudgment of time gaps (C1) The estimation of time gaps (e.g. time left to approaching vehicle, stop sign, traffic lights, etc.) is incorrect. In order to misjudge a time gap to object (e.g. approaching vehicle, stop sign, traffic lights, etc.) must have been observed!

Then, if the crash data supports it, the analysis can continue by looking at the other genotype tables which are denoted with letters B to Q as illustrated in table 32. For instance the "OBSERVATION B" genotype table can be checked to find additional links to other genotypes, and so on.

e. Stop rules

DREAM provides three rules to determine when an analysis is finished. Accordingly, an analysis is completed when at least one of the three following rules is fulfilled:

1. The specific genotypes have the status of terminal events. Therefore, if a specific genotype is the most likely cause of a general consequent, that genotype is chosen and the analysis stops.
2. If there exists no general or specific genotypes that link to the chosen consequent, the analysis stops.
3. If none of the available specific or general phenotypes for the chosen consequent is relevant, given the information available about the accident, the analysis stops.

Analysis process

The first three steps of DREAM concern data collection, accident description and context evaluation. The actual accident analysis starts with the fourth step, in which one specific phenotype is defined for every road user involved in the crash. Once the specific phenotypes are defined, the next step is to identify the corresponding general genotypes (i.e. contributing factors) by looking at the phenotype tables that link phenotypes to genotypes. Subsequently, the genotype tables are used to link genotypes to other genotypes. Finally, the analysis continues until one of the three stop rules applies. Furthermore, the phenotypes and genotypes for every road user involved in the crash are displayed in causation charts.

5.4.1.3 Elements identified from crash accident analysis methods

The description of crashes as four phases and the contributory factors and human driver taxonomies provided by the two methods, were identified as two elements that can be transferred to the analysis of crashes involving vehicle automation.

Crash phases

The two traditional crash analysis methods describe crashes as four sequential phases: the driving phase, the discontinuity or rupture phase, the emergency phase and the crash phase.

These four phases can be used to help establish the timeline of all crashes including those involving automated driving. Moreover, the rupture phase can facilitate the definition of the starting point of a CAST analysis which considers the failures and unsafe interactions at the physical level that led to the accident. Accordingly, the rupture phase can be defined and examined to assist the identification of the physical failures related to the vehicles (e.g. tire

blowout, mechanical failure, etc.) and the unsafe interactions between several road users and road infrastructure.

Contributory factors and human driver taxonomies

The second safety-specific elements identified from traditional methods were the driver failure taxonomies and contributory factors. The HFF framework proposes a taxonomy for the human driver failures which includes six types of general failures and 20 types of specific failures to the driving activity, and a separate list of explanatory elements (i.e. contributory factors) related to driver, the road, traffic, and the vehicle. Conversely, DREAM provides a classification scheme that combines a driver failure taxonomy and a set of contributory factors related to the driver, technology and organizations.

Table 35 displays the categories for driver failures proposed in the two traditional methods. The information acquisition, diagnosis and prediction categories in the HFF framework correspond to the observation, interpretation and planning categories in DREAM. Although these driver failures categories are very useful for the analysis of the driver's role in today's crashes, they are focused on the driver and today's driving task; it is unclear whether or not these categories can capture the human failures in the context of automation and the failures related to automaton.

One solution could have been to go back to the conceptual foundation of the driver failure taxonomies (i.e. the generic models of human information processing and COCOM) to define human failures related to the interactions with automation and automation failures. Nevertheless, the control flaws classification provided in STAMP already offers a categories of things that can go wrong with human drivers and automation. Therefore, it was determined that instead of extending the driver failure taxonomies from traditional methods, the control flaws classification provided in STAMP was going to be applied on an automated driving system in order to create taxonomies for the human and automated controllers which are specific to automated driving.

Table 35 – Driver failure taxonomies in the HFF framework and DREAM

Driver failure taxonomies	
HFF	**DREAM**
Information acquisition	**Observation**
Failure in detection linked to lack of visibility	Missed observation
Focalized acquisition of information	Late observation
Hasty search of information	False observation
Interruption in information acquisition	
Failure to recognize acquisition demands	
Diagnosis	**Interpretation**
Incorrect evaluation of road difficulty	Misjudgment of the situation
Incorrect evaluation of gap	Misjudgment of time gaps
Incorrect understanding of site configuration	Incomplete judgment of the situation
Incorrect understanding of a maneuver undertaken by another road user	
Prediction	**Planning**
Expectance of another road user maneuver	Priority error
Expecting adjustments by another road user	
Expecting no perturbations ahead	
Decision-making	
Directed violation	
Deliberated violation	
Violation-error	
Action	
Vehicle controllability	
Guidance defect	
Overall failure	
Loss of psycho-physiological capabilities	
Alteration of sensimotor and cognitive capabilities	
Overstretching cognitive capacities	

The contributory factors associated to vehicle crashes proposed by the two traditional methods were harmonized and grouped in order to provide an overview of all the factors. As see in figure 39, the contributory factors from the two methods were classified into four main categories: the human driver, the vehicle, infrastructure and traffic. Moreover, each category also includes sub-categories and the indication of the method source; for instance the human driver category is sub-divided into state of the driver, driver experience, task performance and organization.

The overview of contributory factors can be incorporated into CASCAD, to provide guidance regarding the causal factors found in today's crashes. While many of these factors will still be relevant for manual driving and in some cases for automated driving (e.g. a tire blowout and a degraded state of the road), the new causal factors related to vehicle automation and to its side effects on manual driving, are missing.

Human Driver		
State of the driver	**HFF**	**DREAM**
Drowsiness	X	X
Malaise	X	X
Under the influence	X	X
Inattention	X	X
Stress, anger, fear	X	X
Long reaction time	X	
Hypo vigilance	X	
Distraction	X	X
Permanent impairment	X	X
Driver experience	**HFF**	**DREAM**
Automatic driving	X	
Expectance of certain site configuration		X
Novice driver	X	X
Episodic driving	X	
Unknown site	X	
Insufficient situation experience	X	
Unknown vehicle	X	
Task performance	**HFF**	**DREAM**
Priority feeling	X	
Over trust in other road user's signals	X	
Time constraints	X	X
Banalization of the situation	X	
Speeding	X	
Risky driving	X	X
Direction problem	X	
Secondary non-driving activity	X	
Organization	**HFF**	**DREAM**
Irregular working hours		X
Heavy physical activity		X
Inadequate training		X

Vehicle		
Vehicle	**HFF**	**DREAM**
Vehicle size	X	
Tire blowout, pressure	X	X
Mechanical failure	X	X
Lighting failures	X	X
Loading	X	
Cold tires	X	
Tire conditions	X	
Suspension	X	
Temporary HMI problems	**HFF**	**DREAM**
Illumination problems		X
Sound problems		X
Sight obstruction		X
Access limitations		X
Incorrect ITS info		X
Permanent HMI problems	**HFF**	**DREAM**
Illumination problems		X
Sound problems		X
Sight problems		X
Vehicle design and maint.	**HFF**	**DREAM**
Inadequate design of driver environment		X
Inadequate design of communication devices		X
Inadequate construction of vehicle parts/ structures		X
Unpredictable system characteristics		X
Inadequate vehicle maintenance		X

Infrastructure		
Road signs	**HFF**	**DREAM**
Inadequate road signs	X	X
Insufficient advanced road sign information	X	
Road sign design		X
Roads	**HFF**	**DREAM**
Narrow roads	X	
State of the roads	X	
Reduced friction	X	X
Road design	X	X
Road maintenance		X
Infrastructure	**HFF**	**DREAM**
Limited visibility due to infrastructure	X	
Inadequate road lighting	X	
Inadequate infra.	X	
Site complexity	X	

Traffic		
Traffic env.	**HFF**	**DREAM**
Difficulty to obtain an insertion gap	X	
Temporary obstruction of view	X	X
Permanent obstruction of view	X	X
Obstacle on the road	X	X
Absence of road user's signals	X	
Ambiguous road user's signals	X	
Other road user's behavior	X	
Env. disturbance	X	
Weather conditions	**HFF**	**DREAM**
Reduced visibility	X	X
Strong wind	X	X
Communication	**HFF**	**DREAM**
Inadequate transmission to others		X
Inadequate transmission form road environment		X

Figure 39 - Overview of contributory factors

Studies from a human factor perspective have identified issues related to automated driving which may constitute new causal factors, such as the loss of situation awareness, mode confusion, distraction, long-term adaptation and skill degradation, overreliance, motion sickness and communication between the driver and automation (N. Merat and Jamson 2009; Jamson et al. 2013; Natasha Merat and de Waard 2014; NHTSA 2014; Kyriakidis et al. 2017). However, most of these issues have been determined through lessons from other automated domains such as aviation and studies on driving simulators. The empirical validation of such issues on real-driving environments and long-term use has not yet been demonstrated.

5.4.2 Elements to facilitate the application of CAST on automated driving

While the previous sub-section identified two elements from traditional methods which can be incorporated into CAST to assist the analysis of crashes involving automated driving systems, there is also a need to provide elements that assist the application of CAST on automated driving. This sub-section introduces three guidance elements which were developed to this end:

1. Control structure at the physical level.
2. Control flaws classification for the human driver controller and the automated controller.
3. Control structure of the entire road transport system containing indirect controllers.

5.4.2.1 Control structure at the physical level

A generic control structure (illustrated in figure 40) was built to facilitate the identification of failures and unsafe interactions at the physical level. The control structure displays the interactions in terms of control actions (black arrows) and feedbacks (blue dotted arrow) between two vehicles and the infrastructure, and the influence of external factors (e.g. weather). The control structure at the physical level allows the identification of unsafe interactions in terms of the vehicles involved. For instance, vehicle A did not stop at an intersection or vehicle B made a left turn on a forbidden turn. Additionally, the control structure in figure 40 can be modified to consider crashes that only involve one vehicle, such as run-off-road collisions by eliminating vehicle B and focusing on the interactions between

vehicle A and infrastructure, or modified to replace vehicle B by a pedestrian for crashes involving a pedestrian.

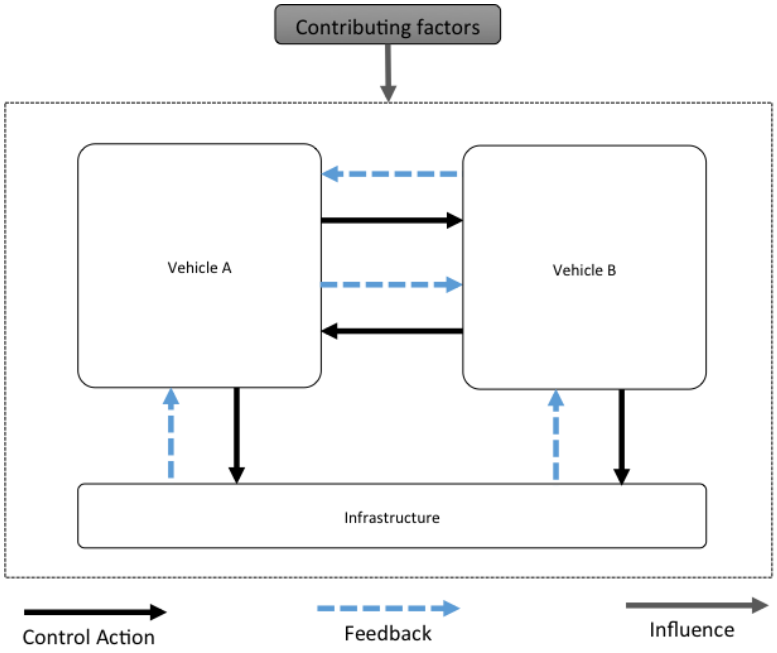


Figure 40 - Generic control structure of interactions at the physical level

5.4.2.2 Control flaws classification

Control flaws classifications for the human driver and for automation were developed to assist the analysis of direct controllers. The classifications were established by analyzing the control structure displayed in figure 41, which includes the direct controllers of an automated vehicle (i.e. Vehicle A) and a non-automated vehicle (i.e. Vehicle B). Vehicle A represents a vehicle equipped with a generic automated driving system with two direct controllers: the automated controller and the human driver controller. In contrast, Vehicle B represents a non-automated vehicle with a human driver controller as its only regulator. The interactions between the direct controllers, vehicle A, vehicle B, and the infrastructure are illustrated in terms of feedback loops (blue arrows) and control actions (black arrows).

The two human driver controllers receive feedback on other road users and on the environment through human perception, which helps them determine what control actions are needed to control the motion of the vehicle and keep it in a safe state. Additionally, the human driver can engage the automated driving system and receive feedback from automation via the

HMI. The automated controller receives feedback on the driving environment and the driver via the vehicle sensors and networks to operate the vehicle. Lastly, there are external and contributory factors related to all the components of the system that have an influence on the control structure, such as the driver's experience and physiological state, vehicle maintenance, the state of the roads, the lighting and weather conditions, etc.

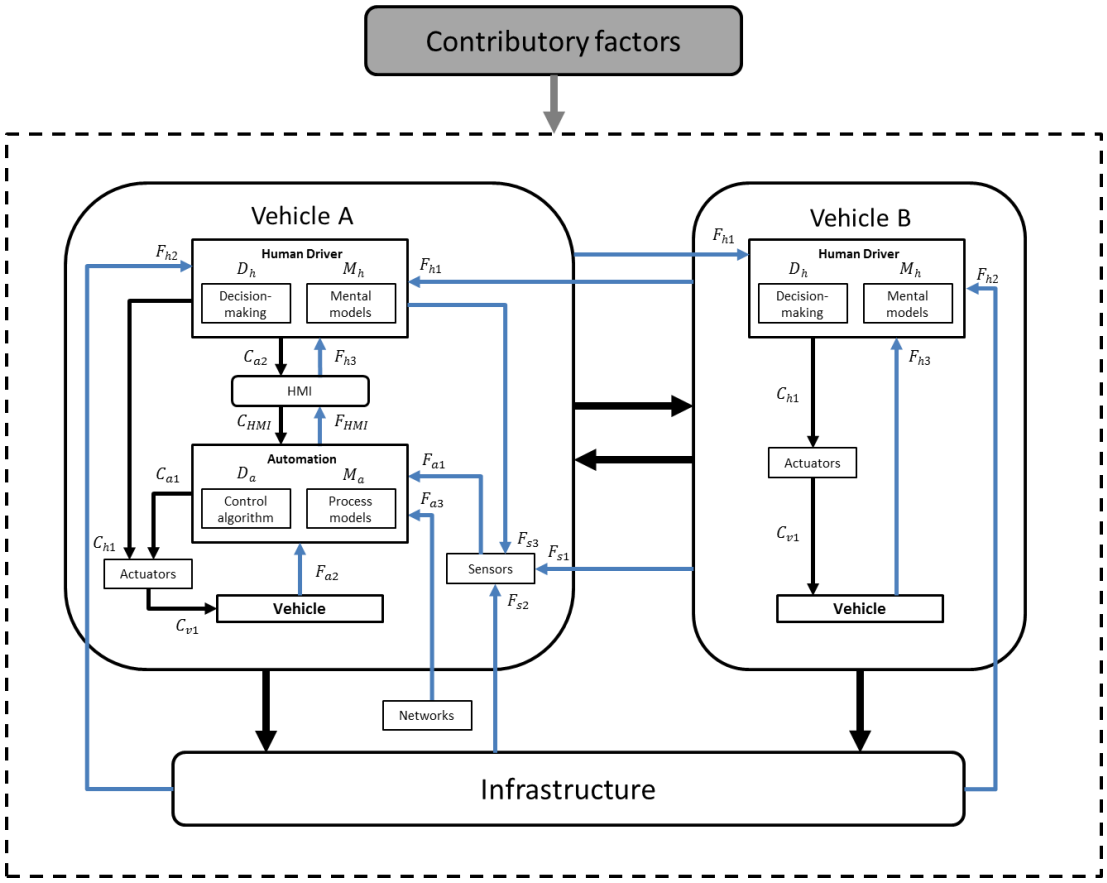


Figure 41 - Control structure of the direct controllers

All the feedback loops, control actions, and components of the control structure were analyzed using the general control flaw classification (figure 19) provided by Leveson (Leveson 2011; Leveson and Thomas 2013). The analysis identified 38 control flaws associated to the human driver controller and 36 control flaws associated to the automated controller.

For example, the analysis of the feedback loop F_{h1} related to the feedback that the human driver receives from other road users, led to the identification of a control flaw in which another road user provides incorrect feedback e.g. another vehicle turns on the left indicator

before turning right. Further, the analysis of the control action C_{a1} corresponding to the actions that automation provide to execute lateral and longitudinal control of the vehicle, enabled the identification of a control flaw in which the control actions are ineffective e.g. automation does not provide enough braking to avoid a rear-end collision with another vehicle.

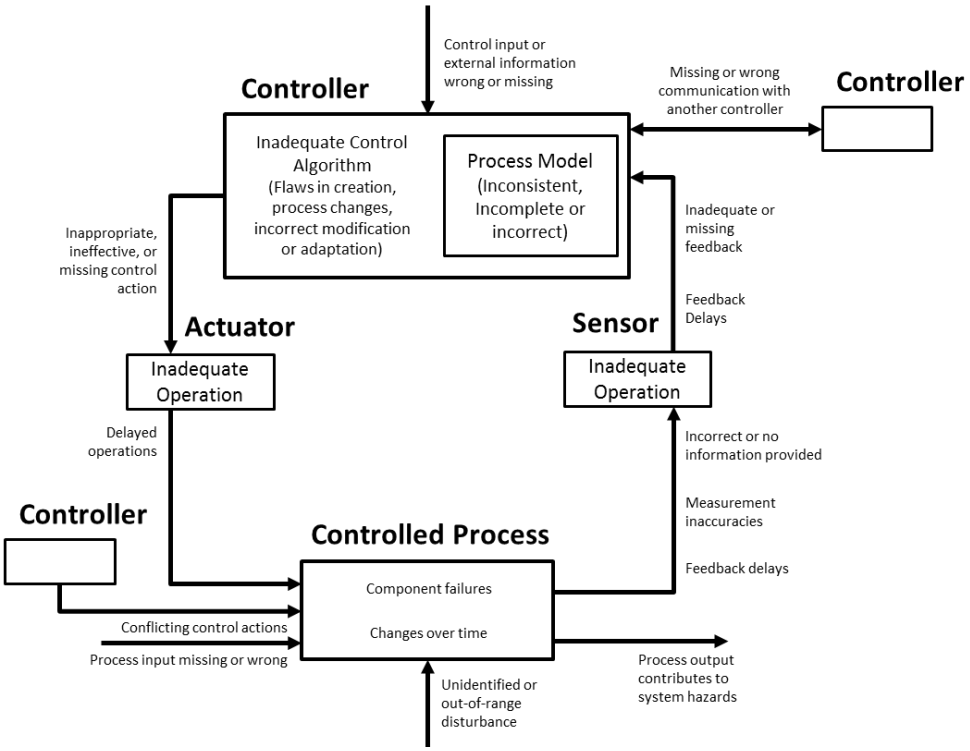
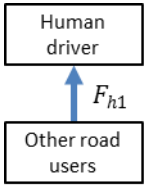
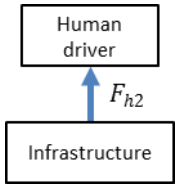
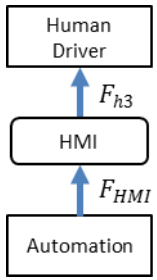
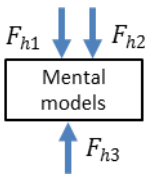


Figure 19 – Potential control flaws related to the control loop

The control flaws identified for the human driver and automation were organized into two tables (table 36 and table 37 respectively) according to four categories: perception, process models, decision-making and action execution. Additionally, examples were defined to illustrate the control flaws and the SAE level concerned by the examples (SAE International 2016).

Table 36 - Control flaws for the human driver controller

Human Driver Controller							
Category		**Control flaw**	**Example**	**SAE level**			
				0	**1-2**	**3**	**4**
Perception		Incorrect information provided by another road user	Another road user provides incorrect information when he/she turns on the left indicator before turning right	X	X	X	
		No information provided by another road user	Another road user does not provide information when he/she does not turn on the left indicator light before turning left	X	X	X	
		Inadequate human perception of feedback on another road user	The human driver inadequately perceives that a biker stopped before an intersection when the biker has not stopped	X	X	X	
		Missing human perception of feedback on another road user	The human driver does not perceive another road user on the adjacent lane of a highway	X	X	X	
		Human perception delays of feedback on another road user	The human driver perceives/sees another vehicle on the adjacent lane of a highway too late	X	X	X	
		Inadequate feedback due to inadequate operation of digital infrastructure	A variable message sign indicating work zone/accident ahead does not display the message due to inadequate screen operation	X	X	X	
		Incorrect information provided by infrastructure	The road has single broken white line markings on a road section in which overtaking is not allowed	X	X	X	
		No information provided by infrastructure	There is no traffic sign indicating a “no right turn” sign on an area in which it is prohibited to turn right	X	X	X	
		Inadequate human perception of infrastructure feedback	The human driver inadequately perceives that the traffic light he is concerned with is green because he/she looks at the wrong traffic light	X	X	X	
		Human missing perception of infrastructure feedback	The human driver does not perceive/see a “no U turn” sign before doing the U turn	X	X	X	
		Human perception delays of infrastructure feedback	The human driver perceives a stop sign too late	X	X	X	
		Incorrect information provided by the automated controller (F_{HMI})	The automated controller provides the HMI incorrect information relative to the speed of the vehicle		X	X	X
			The automated controller provides incorrect information to the HMI on the availability of AD mode			X	X
		No information provided by the automated controller (F_{HMI})	The automated controller does not provide information to the HMI on the non-detection of lane markings		X	X	X
			The automated controller does not provide takeover request to the HMI			X	X
		Feedback delays of information provided by the automated controller (F_{HMI})	The automated controller provides the HMI information on reaching the minimum speed distance with delays		X	X	X
			The automated controller provides takeover request to the HMI with delays			X	X
		Inadequate feedback (F_{h3}) due to inadequate operation of the HMI	The automated controller detects that the minimal safety distance is reached and sends a signal to trigger an alarm but the HMI does not display the alarm due to inadequate component operation (e.g. component failures)		X	X	X
		Inadequate human perception of HMI feedback (F_{h3})	The human driver inadequately perceives the feedback on ADAS deactivation coming from the HMI		X		
			The human driver inadequately perceives the takeover request on the HMI			X	X
		Missing human perception of HMI feedback (F_{h3})	The human driver does not perceive the ADAS deactivation warnings coming from the HMI		X		
			The human driver does not perceive the takeover request coming from the HMI			X	X
		Human perception delays of HMI feedback (F_{h3})	The human driver perceives the feedback from the HMI too late		X	X	X
			The human driver perceives the takeover request from the HMI too late			X	X
Mental models		Inadequate mental model of the traffic situation (M_h)	The human driver believes that he does not need to stop at an intersection because he/she has the right of way	X	X	X	
			The human driver believes that the traffic situation has insertion gaps which enable him/her to change lanes because s/he misjudges the size of the gaps (the gaps are not large enough to allow the insertion)	X	X	X	
		Inadequate mental model of other road users’ behaviors (M_h)	The human driver believes that the traffic situation has insertion gaps which enable him/her to change lanes due to the absence of blind spot detection feedback		X	X	

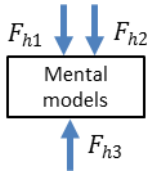
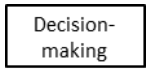
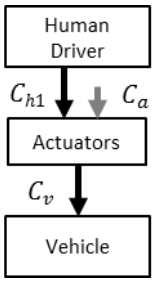
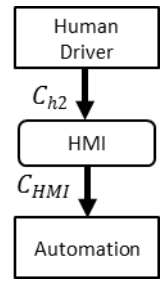
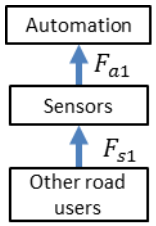
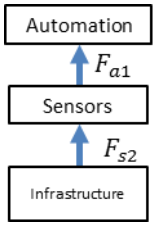
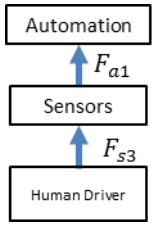
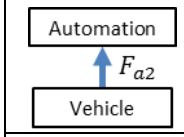
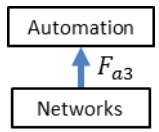
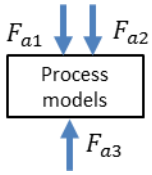
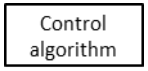
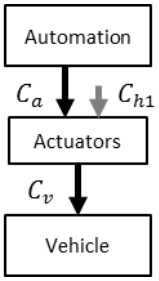
Mental Models		Inadequate mental model of other road user's behavior (M_h)	The human driver believes that another road user is driving at an adequate speed based on human perception	X	X	X	
			The human driver thinks that another road user has a safe behavior based on the driving environment view provided by the HMI			X	X
		Inadequate mental model of other road user's intentions (M_h)	The human driver thinks that another road user will allow him to change lanes because he/she sees the other road user reduce speed	X	X	X	
			The human driver thinks that the other road user will respect traffic rules			X	X
		Inadequate mental model of automation (M_h)	The human driver is unaware of the ADAS operational domain		X		
			The human driver does not know how to operate the automated driving system (e.g. unaware of engagement commands and sequences, takeover validation procedure, etc.)			X	X
		Inadequate mental model of the driving mode (M_h)	The human driver does not understand the change of interface between manual driving to automated driving mode; s/he believes that the vehicle is still on manual driving mode		X	X	X
		Inadequate mental model of driving mode transitions (M_h)	The human driver has an inadequate mental model of the driving mode transitions and the procedures to hand over control to automation and take over control from automation			X	X
			The human driver perceives the request but does not know/understand that it is a takeover request			X	X
Decision-making		Inadequate human driver decision-making (D_h)	The human driver deliberately makes a decision that leads to the violation of a traffic rule	X	X	X	
			The human driver deliberately makes a decision to engage automation outside of the design limits			X	X
			The human driver unintentionally makes a decision that leads to the violation of a traffic rule			X	X
			The human driver unintentionally makes a decision to engage automation outside of the design limits		X	X	X
			The human driver makes a decision that leads to an unsafe behavior	X	X	X	X
Control Actions		Inappropriate control action (C_{h1})	The human driver provides acceleration when the distance to a vehicle in the front is less than the safety distance	X	X		
		Ineffective control action (C_{h1})	The human driver does not brake hard enough to stop at an intersection				
		Missing control action (C_{h1})	The human driver does not provide braking when the distance to a vehicle in the front vehicle is less than the safety distance	X	X		
			The human driver does not provide override when the automation puts the vehicle in an unsafe situation		X	X	X
		Delayed control action (C_{h1})	The human driver provides braking too late when the distance to a vehicle in the front is less than the safety distance	X	X		
		Inadequate actuator operation (C_v)	The human driver provides adequate steering but the power/ hydraulic actuators of the steering system operate inadequately	X	X		
		Delayed actuator operation (C_v)	The human driver provides adequate steering but the power/ hydraulic actuators of the steering have a delayed operation	X	X		
		Conflicting control action (C_{h1} and C_a)	The human driver provides braking that conflicts with braking actions provided by automation		X	X	X
		Inappropriate control action (C_{h2})	The human driver activates an ACC on urban roads		X	X	X
			The human driver validates takeover request when he is not ready to resume the manual driving mode			X	X
		Missing control action (C_{h2})	The human driver does not validate AD mode activation and releases control over the vehicle			X	X
			The human driver does not validate the takeover request			X	X
		Delayed control action (C_{h1})	The human driver validates takeover request too late			X	X
		Inadequate actuator operation (C_{HMI})	The human driver pushes the AD activation button via the HMI but the button has a component failure			X	X
		Delayed actuator operation (C_{HMI})	The human driver pushes the button to validate takeover request but there is a delay in the button operation			X	X
		Conflicting control action (C_{h2} and C_{h1})	The human driver provides ADAS activation and provides acceleration at the same time		X		
			The human driver provides AD mode activation and provides acceleration at the same time			X	X

Table 37 - Control flaws for the automated controller

Automated Controller							
Category		**Control flaw**	**Example**	**SAE level**			
				0	**1-2**	**3**	**4**
Perception		Measurement inaccuracies on road users feedback measured by sensors (F_{s1})	Radar provides inaccurate measurements on the distance to a vehicle or an obstacle		X	X	X
		Sensor inadequate operation (F_{a1})	Radar has an inadequate operation due to a component failure		X	X	X
		Inadequate/incorrect feedback on another road user provided by sensors (F_{a1})	Camera provides inadequate/incorrect feedback on the presence of another road user (e.g. detecting a tree as a pedestrian)		X	X	X
		Feedback delays on other road users feedback provided by sensors (F_{a1})	Radar provides feedback on distance to another vehicle with a delay of TBD ms		X	X	X
		Missing feedback on road users/obstacles provided by sensors (F_{a1})	Radar does not detect the presence of another road user		X	X	X
		Incorrect/no information provided by infrastructure (F_{s2})	Camera adequately detects a “left turn allowed” traffic sign in an area where it is unsafe to allow left turns		X	X	X
			The road has degraded road markings and consequently the camera cannot detect road markings			X	X
		Measurement inaccuracies on infrastructure feedback measured by sensors (F_{s2})	Camera provides inaccurate measurements on road markings		X	X	X
		Sensor inadequate operation (F_{a1})	Radar does not detect an obstacle on the road due to the radar’s inadequate operation		X	X	X
		Inadequate feedback on infrastructure provided by sensors (F_{a1})	Camera detects an obstacle and classifies it as a tree but the obstacle is in fact is a pedestrian		X	X	X
		Missing feedback on infrastructure provided by sensors (F_{a1})	Radar does not detect an obstacle on the road (although it was operating adequately)		X	X	X
		Feedback delays on infrastructure feedback provided by sensors (F_{a1})	Camera provides feedback on front vehicle detection with delays		X	X	X
		Incorrect information provided by the human driver (F_{s3})	The human driver tapes a soda can to the wheel to fool automation into thinking s/he has hands on the wheel		X	X	X
		Sensor inadequate operation (F_{s3})	The “hands on steering wheel” sensor does not detect the presence of the driver’s hands due to inadequate operation		X	X	X
		Inadequate feedback on human driver provided by sensors (F_{a1})	The driver monitoring sensors provide feedback that the driver is in the control loop because it detects that the driver has hands on the steering wheel/ is looking at the road		X	X	X
		Missing feedback on human driver provided by sensors (F_{a1})	The “feet on pedals” sensors do not provide feedback		X	X	X
		Feedback delays on human driver feedback provided by sensors (F_{a1})	The “feet on pedals” sensors provide feedback with delays		X	X	X
		Incorrect information provided by the vehicle (F_{a2})	The cloud network provides information about a working zone with delays		X	X	X
		No information provided by the vehicle (F_{a2})	The CAN bus provides incorrect information about the position of the throttle valve		X	X	X
		Feedback delays of information provided by the vehicle (F_{a2})	The CAN bus provides information about the position of the throttle valve with delays		X	X	X
		Incorrect information provided by networks (F_{a3})	The V2I network provides incorrect information to automation about the speed limit		X	X	X
			The cloud network provides incorrect information on AD road section verification			X	X
		No information provided by networks (F_{a3})	The cloud network provides no information about a weather conditions		X	X	X
		Feedback delays of information provided by networks (F_{a3})	The cloud network provides information about a working zone with delays			X	X

Process models		Inadequate model of the traffic situation (M_a)	The automated controller believes that there is heavy traffic when traffic is fluid		X	X	X
			The automated controller does not realize that there is an ambulance and that it needs to emergency vehicles			X	X
			The automated controller is unaware that it is committing a traffic violation				X
		Inadequate model of other road user's behavior (M_a)	The automated controller is unaware that a road user is changing lanes			X	X
			The automated controller is unaware that a road user is having a reckless behavior			X	X
		Inadequate model of other road user's intentions (M_a)	The automated controller thinks that another road user will allow it to change lanes			X	X
		Inadequate model of automation (M_a)	The automated controller is unaware that the vehicle is outside of automation's operational design domain			X	X
			The automated controller has an inconsistent model of the driving mode status, it believes that the automated driving system is engaged when it is not			X	X
		Inadequate mental model of driving mode transitions (M_a)	The automated controller is unaware that a takeover request is necessary			X	X
			The automated controller thinks that it is safe to enter AD mode			X	X
		Inadequate model of the human driver (M_a)	The automated controller believes that the driver is ready to regain the control of the vehicle			X	X
Decision-making		Inadequate control algorithm (D_a)	The control algorithm generates a command that violates traffic regulations		X	X	X
			The control algorithm provides actions when the vehicle is in manual driving mode		X	X	X
			The control algorithm puts the vehicle in an unsafe situation due to adaptation (machine learning)			X	X
			The control algorithm puts the vehicle in an unsafe situation due to incorrect modifications (software upgrades)		X	X	X
Control Actions		Inappropriate control action (C_a)	The automated controller provides acceleration when the distance to a vehicle in the front is less than the safety distance		X	X	X
			The automated controller provides acceleration when the vehicle is on manual driving mode		X	X	X
		Ineffective control action (C_a)	The automated controller does not brake hard enough to stop when the distance to a vehicle in front is less than the safety distance		X	X	X
		Missing control action (C_a)	The automated controller does not provide braking when the distance to a vehicle in the front vehicle is less than the safety distance		X	X	X
			The automated controller does not provide the minimal risk maneuver when the driver does not respond to the takeover request			X	X
		Delayed control action (C_a)	The automated controller provides braking too late when the distance to a vehicle in the front is less than the safety distance		X	X	X
		Inadequate actuator operation (C_a)	The automated controller provides adequate steering but the power/ hydraulic actuators of the steering system operate inadequately		X	X	X
		Delayed actuator operation (C_v)	The automated controller provides adequate steering but the power/ hydraulic actuators of the steering have a delayed operation		X	X	X
		Conflicting control action (C_a and C_{h1})	The automated controller provides a control action that conflicts with a control action provided by the human driver		X	X	X

5.4.2.3 Control structure of the entire road transport system

The last guidance element is a control structure of the entire road transport sociotechnical system in France (figure 42) which was built to assist the analysis of indirect controllers at the higher levels of the system. The control structure models the road transport system according to six hierarchical levels. In the first three levels, the controllers that influence development and operation of the transport system are the same. In turn, levels four and five display a distinction between the development and operation of the system. Additionally, the lowest level (level six) represents the operating process with the basic components: road users, vehicles and infrastructure. Throughout the structure the feedback loops between the levels are displayed with upward blue dashed arrows from lower to higher levels, and the safety constraints (which are imposed through control actions) are illustrated with downward black arrows from higher level to lower levels. Lastly, the grey arrow indicates the environment conditions (weather, lighting, etc.) and disturbances that contribute to crashes.

Level 1 International and European stakeholders

The international and European stakeholders at level one influence the French transport system via International and European policies, ratings, standards and pressure, and receive feedback from the French transport system regarding the status of road safety, the impact of existing measures, and new measures, in meetings, reports, and studies.

The United Nations (UN) and the European Union (EU) establish legally binding policies to regulate the development and operation of the road transport system; for instance, the UN working party on Road Traffic Safety (WP.1) serves as a guardian of the policies aimed at harmonizing traffic rules (e.g. the 1949 Geneva Convention and the 1968 Geneva Convention), and the UN World Forum for Harmonization of Vehicle regulations (WP. 29) defines technical regulations for vehicles. Moreover, the EU disposes of two types of legally-binding policies:

1. The EU regulations which represent directly applicable legislation within the member states.
2. The EU directives which must be implemented into national legislations (e.g. the directive on type-approval for the safety and environmental requirements a vehicle must comply with before being place on the EU market).

Furthermore, the UN and the EU also provide non-legally binding policies; such as best practices, white papers, action programs, funding, etc.

The International Road Assessment Program (iRAP), the European Road Assessment Program (EuroRAP), the Global New Car Assessment Program NCAP (Global NCAP) and the European New Car Assessment Program NCAP (Euro NCAP), influence the transport system via road and vehicle ratings which encourage the development of safer roads and safer vehicles. International and European standard bodies like the International Organization for Standardization (ISO), the European Committee for Standardization (CEN), and the Society of Automotive Engineers (SAE), influence the road transport system through standards on vehicle design, transport information and control systems, road traffic system management, etc. Lastly, the international and European associations (including industry and road safety associations) attempt to influence the policies established by the EU and UN via lobbying and pressures.

Level 2 National parliament and legislatures

The level two influences the lower levels of the structure via legislation and receives feedback such as law proposals and reports during meetings and sessions. However, in France, most of the control mechanisms for the safety of the road transport system are enforced by policies established at the government level called *"projet de loi"*.

Level 3 Government agencies, courts, research bodies, standard bodies, associations, etc.

The controllers at level three influence the lower levels (levels four to six) of the road transport development and operations system through control mechanisms like national policies, legal penalties, innovations and pressure. Moreover, level three receives feedback from levels four to six in terms of accidents and incidents, and reports.

Several government agencies intervene in the definition of the national policies regarding road transport. The Ministry of the interior is responsible for the definition of road safety and road safety education policies. Moreover, they also set guidelines for the road intervention of the national police, and standards for driving education and certification. The Ministry of Ecology is in charge of enforcing the vehicle technical regulations (which derive from EU directives) to

develop and manage the national road network (i.e. highways and national roads), to establish guidelines for the other types of road network owned by departments and communes, and to define policies for road infrastructure safety. The Ministry of justice prepares and enforces law projects on criminal justice, civil justice and administrative justice. The Ministry of Health sets the regulatory framework for the management of persons injured on the roads. The Ministry of Labor establishes policies to reduce road risk as an occupational risk (road crashes are the major cause of mortality at work in France). The Ministry of Education is in charge of the first phase of road safety education which starts at school. Finally, Courts are in charge of interpreting and applying the law to carry out administration of justice in the criminal, civil and administrative regimes.

Public and private research bodies provide feedback to the government on the state of the road transport safety, and the impact assessment of safety measures. Additionally, they develop technologies that aim at improving road safety (i.e. a new vehicle system, infrastructure component, etc.). National associations oversee the government's actions and put pressure to influence the policies established by the government via lobbying and pressures.

Level 4 Company (development)

The controllers at the left side of level 4, represent the companies and organizations that design and develop the basic components of the operating process: infrastructure, vehicles and road users. Organizations such as local governments, road infrastructure companies and network companies, are in charge of the infrastructure components, automakers and suppliers are in charge of the vehicle component, and organizations, such as driving schools and road safety education centers, are responsible for the training of road users.

These controllers receive feedback from the lower from the operating process regarding crashes and incidents. Finally, these controllers also provide guides and instructions for the level five (maintenance), for example automakers give garages instructions on how to access/repair certain vehicles.

Level 4 Company (operations)

The controllers at the right side of level four represent the companies and organizations that participate in the operation of the system. This level covers organizations such as companies with professional drivers, car rental companies and car dealerships, which provide with vehicles to drivers. Additionally, it includes organizations that operate road infrastructure such as local governments and road infrastructure companies. Furthermore, it comprises the organizations that train operators who are in direct contact with road users in the operating process, like the police and the emergency services. This level receives feedback from the operating process like crashes, complaints, traffic data, etc. Some of the controllers in level four also play a role in level five; for instance, the road infrastructure companies also perform the maintenance of road infrastructure.

Level 5 Maintenance

Level five (maintenance) contains the controllers that actively participate in the maintenance of the operating process' components, for instance garages, road maintenance operators and the companies that perform software updates. It receives feedback regarding vehicle requests for a repair, incident reports on infrastructure, and software upgrade requests.

Level 6 Operating process

The last level is influenced by all the safety constraints imposed at the higher levels of the structure. It involves human driver, vehicles and the driving environment (i.e. infrastructure, other road users, etc.); additionally, this level is influenced by environment conditions (weather, lighting, etc.) and disturbances that contribute to crashes. This level contains the control structure at the physical level and the direct controllers of the system i.e. the human driver controller and the automated controllers.

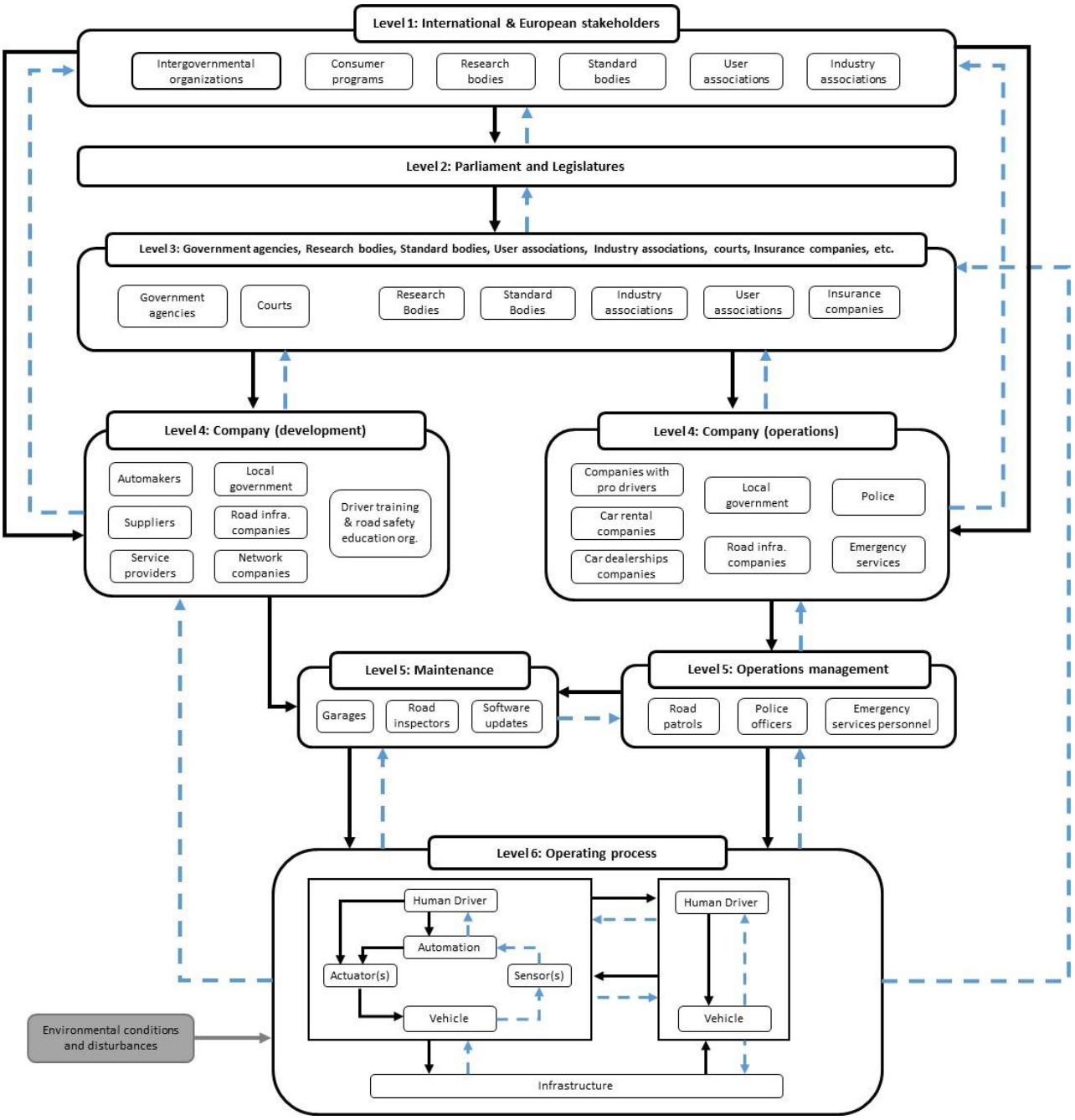


Figure 42 - Generic control structure of the road transport system

5.4.3 CASCAD

This sub-section introduces the CASCAD method in two stages; firstly, it describes how the identified road safety-specific elements and the newly developed guidance elements were incorporated into CAST in order to create CASCAD. Secondly, it illustrates the application of CASCAD using the information available on the widely publicized crash involving a Tesla in May 2016.

5.4.3.1 CASCAD process

CASCAD uses five main steps described in CAST as its backbone, and incorporates the two road safety-specific elements identified from existing crash analysis methods and the three elements that were developed to facilitate the application of CAST to automated driving. As illustrated in figure 43, the first step of CASCAD remains the same as in CAST; it consists of establishing the system engineering foundation by defining the accidents, hazards and constraints of the system and by building the control structure. In the second step, the description of crashes as four phases and the control structure of the physical interactions are included to facilitate the identification of failures and unsafe interactions at the physical level.

For the analysis of the direct controllers, the control flaws classification and the contributory factors were incorporated to assist the understanding of why the human driver and automated controllers behaved unsafely. Further, the control structure of the road transport system in France (which can easily be adapted to fit the transport system in other regions) was included in the fourth step to assist the analysis of indirect controllers. The last step of the CASCAD analysis is the same as in CAST i.e. generating recommendations that aim at redesigning the system and preventing accidents based on the results of the analysis.

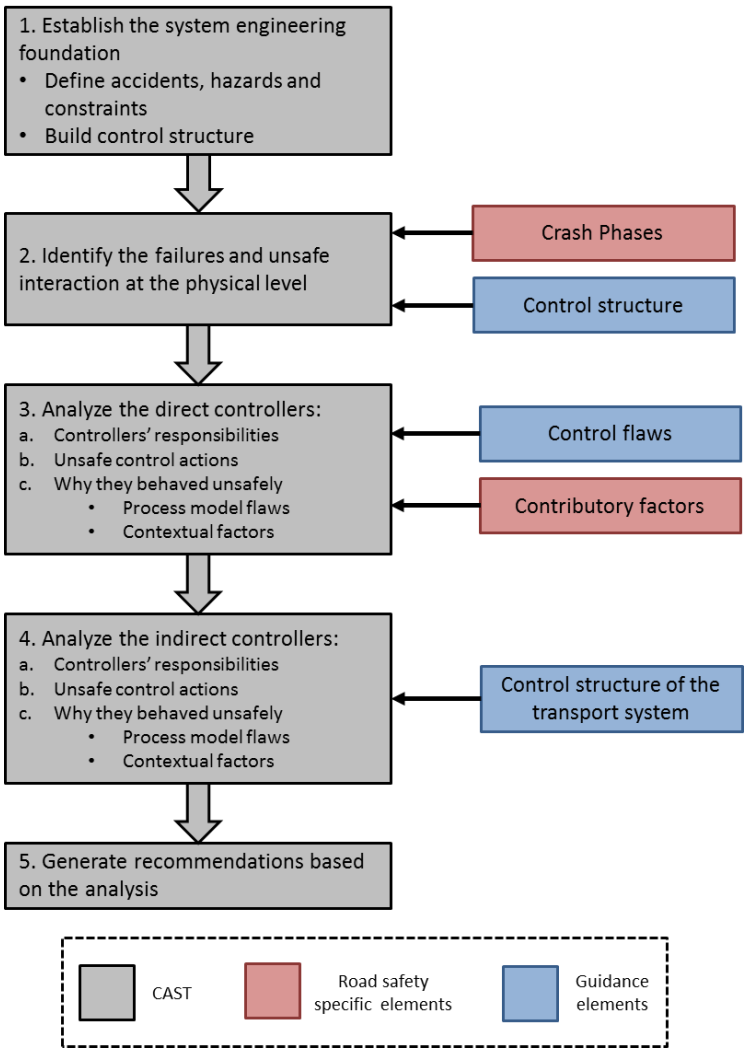


Figure 43 – CASCAD process

5.4.3.2 CASCAD illustration using the Tesla crash

The Tesla crash considered for the illustration of CASCAD is described using the available information on the internet and on the NTHSA crash reports. Subsequently a CASCAD analysis comprising the five steps described in figure 43 is conducted on the Tesla crash to show the application of the method.

Description of the Tesla crash

On May 7 2016 at 4:40 pm, a 2015 Tesla Model operated in Autopilot mode travelling on US Highway 27 in Florida, struck a 2014 Freightliner Cascadia truck-tractor in combination with a 16.2 meter semitrailer operated in manual mode (National Transportation Board 2016; Habib

2017). At the time of the crash, it was daylight and the weather was clear and dry. The collision occurred when the 62-year-old truck driver was making a left turn on an uncontrolled intersection, as the Tesla, which had the right of way, approached the intersection at 119 km/h. As illustrated in figure 44, the Tesla hit the trailer at 119km/h with an angle of 90° and then passed underneath the trailer. After exiting from underneath the trailer, the Tesla veered off the road, travelled approximately 90.5 meters and stroke two fences before colliding with a utility pole. The Tesla travelled an additional 15 meters, during which it rotated counterclockwise and finally came to rest. The collision resulted in fatal injures for the only occupant of the Tesla, a 40-year-old male driver.

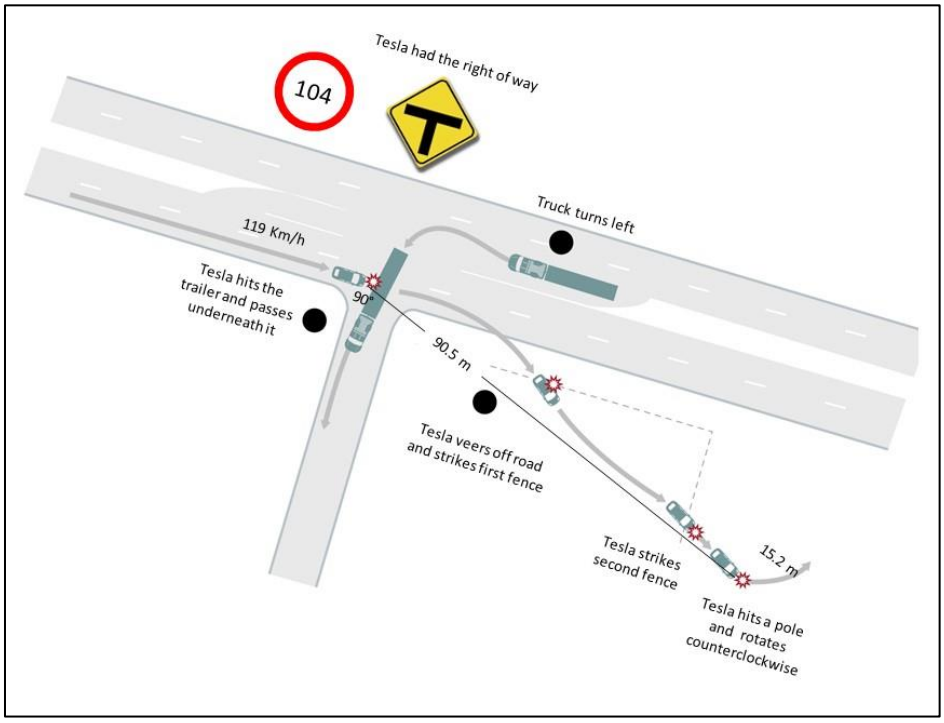


Figure 44 – Tesla crash (Singhvi and Russell 2016)

CASCAD analysis on the Tesla crash

The CASCAD analysis on the Tesla crash is an example to illustrate the application of the five steps of the method on a crash involving automated driving. The analysis was based on the official NHTSA reports on the crash (National Transportation Board 2016; Habib 2017) and on the available information on the internet (Singhvi and Russell 2016; Lambert 2016; The Tesla Team 2016). Although all the necessary information to complete the analysis was not available,

as aforementioned, the intention is to show how the method can be applied rather than to obtain meaningful results from the analysis.

1. Define accidents, hazards and violated constraints

The following accident, hazard and safety constraint were defined for the Tesla crash:

Accident: Human loss due to a vehicle collision.

System hazard: Violation of minimal safety distance between a passenger vehicle and a truck.

Safety constraint: The safety control structure must prevent the violation of minimal distance between a vehicle and a truck.

2. Identify failures and unsafe interactions at the physical level

In the second step of CASCAD, the four phases of the crash were described for the Tesla and the truck in order to set the timeline of the crash. Next, the control structure at the physical level and the rupture phases were considered to identify the physical failures and the unsafe interactions.

Description of the crash:

The four phases of a crash were described for the two vehicles:

Tesla:

Driving phase: The Tesla is travelling on a highway on a Saturday at 4:40 pm.

Rupture phase: The Tesla does not slow down the vehicle as it approaches an uncontrolled intersection.

Emergency phase: The Tesla violates the minimal safety distance to the truck and does not decrease the speed of the vehicle.

Crash phase: The front of the Tesla strikes the trailer of the truck with a 90° angle at 119 km/h, passes underneath the trailer, leaves the road and hits tow fences and a pole before rotating counterclockwise and coming to rest.

Truck:

Driving phase: The truck is travelling on a highway on a Saturday at 4:40 pm to deliver blueberries.

Rupture phase: The truck estimates that it can engage a left turn maneuver.

Emergency phase: The truck engages a left turn maneuver and does not have the time to stop as the Tesla approaches at 119 km/h.

Crash phase: The Tesla collides with the semitrailer of the truck.

Physical failures:

There were no physical failures involved in the crash.

Unsafe interactions:

- The truck made a left turn too soon at an uncontrolled highway intersection when he did not have the right of way.
- The Tesla vehicle did not slow down/stop when the safety distance to a truck was reached.

3. Analyze the direct controllers of the system

The direct controllers of the two vehicles i.e. automation and human driver for the Tesla and human driver for the truck, were analyzed by identifying the unsafe control actions that lead to the crash, the control flaws behind the unsafe control actions and the contextual factors.

Direct controllers of the Tesla:

Automated controller:

Unsafe control actions (UCAs):

Two unsafe control actions were identified for the automated controller:

- **UCA-1:** Automation's autopilot function did not apply brakes when the safety distance to the truck was violated.
- **UCA-2:** The automatic emergency brake (AEB) system[15] did not provide warnings and braking to avoid or mitigate the crash when the crash was imminent.

[15] The SAE classification does not consider the AEB system as an automated driving system thus it is useful to distinguish the AEB system and the Autopilot function.

Control flaws (CFs):

The control flaws related to the automated controller were examined for the two unsafe control actions. As illustrated in Table 38, four control flaws were identified, two from the perception category and two from the mental model category.

- **CF-1:** The camera provided inaccurate measures to automation indicating that there was no obstacle because the camera did not detect the white trailer against the bright sky.
- **CF-2:** The radar provided inadequate feedback to automation regarding the truck because even though it had detected an obstacle, the radar had been configured to tune out data that could indicate overhead road signs in order to avoid false braking events.
- **CF-3:** The autopilot function had an inadequate model of the traffic situation because it was unaware of the presence of the truck due to incorrect feedback provided by the camera and the radar.
- **CF-4:** Automation had an inadequate model of the human driver because it was unaware that the driver was distracted due to the driver monitoring system design in which the driver's engagement is monitored through the interactions with the steering wheel, turn signal, and speed stetting stalk; it does not monitor if drivers have their eyes on the road.

Table 38 – Example of CASCAD analysis of the automated controller

UCA-1: Automation (Autopilot) did not apply brakes when the safety distance to the truck was violated			
Category	**Control Flaw**	**Contributory factors**	**Description**
Perception	CF-1: Measurement inaccuracies on road users feedback provided by sensors	Sunlight and bright sky influence on cameras	Camera provided inaccurate measures due to the white trailer being against bright sky
	CF-2: Inadequate or incorrect feedback provided by sensors	Algorithm strategies to avoid false positives	The radar provided incorrect feedback because it tuned out the data on the truck obstacle to avoid false braking events (overhead traffic signs).
Model of process	CF-3: Inadequate model of the traffic situation	Reliability of the vehicle perception system	Automation (autopilot) was unaware of the presence of the truck due to incorrect feedback
	CF-4: Inadequate model of the human driver	Driver monitoring system	Automation (autopilot) was unaware that the driver was distracted because the driver monitoring system does not detect when drivers have their eyes off the road

Contributory factors:

The contributory factors identified in the automated controller related to vehicle sensors and automation's process model, were not included in the overview of traditional contributory factors to crashes (figure 39). The AEB system uses the same sensors to detect obstacles and

provide warnings and braking to avoid or mitigate collisions, therefore the control flaws identified for the UCA-2 were the same as the three first control flaws identified for the UCA-1.

Context:

Lastly, in terms of context in which decisions were made, it was daylight with clear weather conditions and there were no known problems with the detection of trucks.

Human diver controller:

Unsafe control actions (UCAs):

- **UCA-3:** The Tesla human driver did not override automation and applied brakes when the safety distance to the truck was violated.

Control flaws (CFs):

As shown in table 39, three control flaws and several contributory factors related to human driver controllers were identified for the UCA-3.

- **CF-5:** The driver did not see the truck (missing human perception) because he was distracted in a secondary non-related driving activity (i.e. looking at a DVD player) and did not look at the road.
- **CF-6:** The driver had an inadequate model of the traffic situation because he was unaware of the presence of the truck due to the fact that he had the right of way (he probably expected other road users to stop at the uncontrolled intersection).
- **CF-7:** The driver had an inadequate model of automation because he over relied on automation and believed that automation was able to monitor the traffic environment and assure safe operation.

Contributory factors:

As opposed to the automated driver controller analysis, most of the contributory factors e.g. distraction, secondary non-driving related activity, etc. were included in the overview of traditional contributory factors to crashes. Conversely, the overreliance on automation and the experience with the system are not included in the overview of traditional contributory factors to crashes.

Context:

The driver was a Tesla fan who liked to operate the vehicle in Autopilot mode.

Table 39 – Example of CASCAD analysis for the Tesla human driver

UCA-3: The Tesla human driver did not override automation and apply brakes when the safety distance to the truck was violated.			
Category	**Control Flaw**	**Contributory factors**	**Description**
Perception	CF-5: Missing human perception of feedback on another road user	Distraction Secondary non-driving related activity	The driver did not perceive the truck because he was distracted in a secondary non-driving related activity (i.e. looking at a DVD player[16]) and did not look at the road
Model of process	CF-6: Inadequate model of the traffic situation	Priority feeling	The driver was unaware of the presence of the truck because he knew that he had the right of way and did not perceive the truck
	CF-7: Inadequate model of automation	Overreliance Experience with the system	Driver believed that automation's monitoring was enough for safe operation (overreliance)

Direct controller of the truck:

Human diver controller:

Unsafe control actions (UCAs):

- **UCA-4:** The truck human driver engaged in a left turn in an uncontrolled intersection when the Tesla was approaching at high speed.

Control flaws (CFs):

- **CF-8:** The driver saw the Tesla but had an inadequate model of the traffic situation because he believed that the Tesla was going to slow down and that he could make the left turn.

Contributory factors):

Since the truck driver was operating the vehicle in manual mode, all of the contributory factors identified were a part of the overview of traditional contributory factors in crashes.

Context:

Lastly, in terms of context, the truck driver was making a delivery which may have caused time constraints and stress.

[16] Some sites state that the Tesla driver was watching a Harry Potter movie although this is unofficial.

Table 40 – Example of CASCAD analysis for the truck driver

UCA-4: Truck human driver engaged a left turn in an uncontrolled intersection when the Tesla was approaching at high speed			
Category	**Control Flaw**	**Contributory factors**	**Description**
Model of process	CF-8: Inadequate model of the traffic situation	Misjudgment of time gap Expectance of certain behaviors	The truck driver perceived the Tesla but believed that the Tesla was going to slow down and thought that he could make the left turn.

4. Analyze the indirect controllers of the system

The control structure of the road transport system in France (illustrated in figure 42) was modified to reflect the particularities of the US transport system. As seen in figure 45, the highest level of the system is the congress that provides federal guidelines to regulate the transport system, followed by the federal government agencies including the department of transportation (DOT), the National Highway Traffic Administration (NHTSA) and the Federal Highway Administration (FHWA). The DOT provides resources to the NHTSA and FHWA, in turn, the NHTSA establishes vehicle measures and federal vehicle standards which must be adopted by the automotive industry, and standards for driving education. Additionally, the FHWA provides infrastructure guidelines for federal highways.

At the third level, the state of Florida covers the state government, the Florida Highway Safety and Motor Vehicles (FLHSMV), and Florida department of transportation (FDOT). The FLHSMV implements driving education and grants driving licenses and the FDOT builds infrastructures according to the guidelines established by the FHWA. The fourth level, concerns the automotive industry which must develop vehicles that are compliant with the standards and guidelines defined by the higher levels of the system. Finally, the fifth level covers the operational process in which the Tesla and the truck collided.

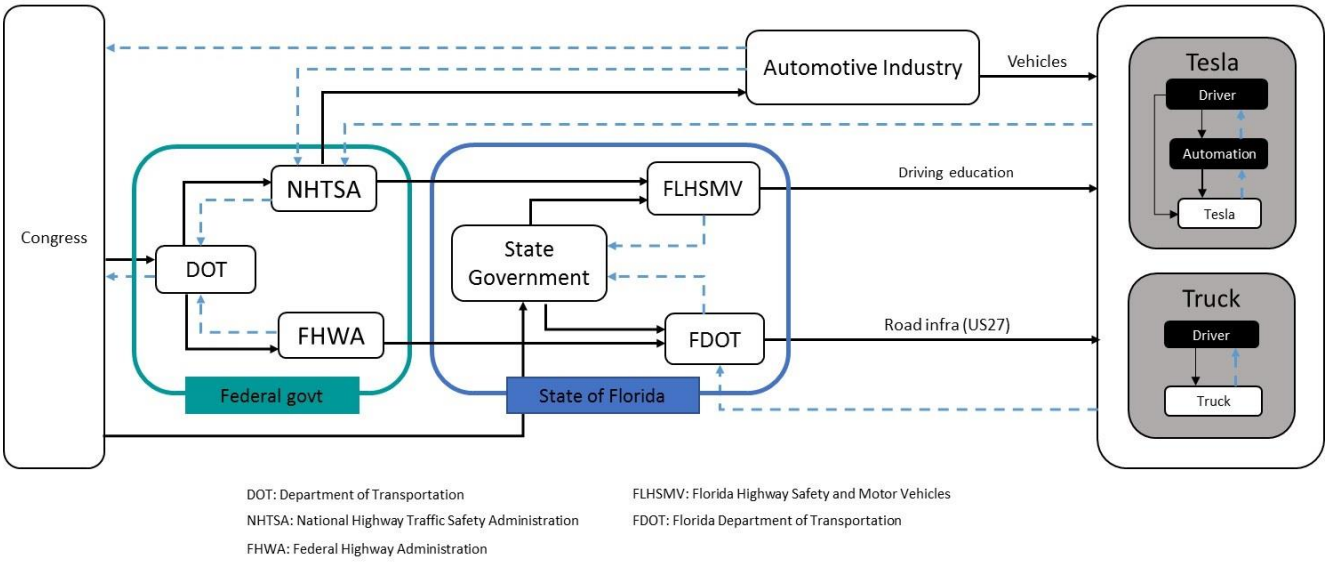


Figure 45 - Control structure of the US transport system

The next step of the analysis, involves analyzing the indirect controllers; CASCAD analyzes indirect controllers by examining their unsafe control actions, model flaws and the context in which controllers made decisions. The information regarding the influence of these controllers on the Tesla accident was not available; however, assumptions were made to establish two examples of the hypothetical analyses of the Tesla Company and NHTSA controllers to illustrate the analysis of indirect controllers. As opposed to direct controllers, there are no control flaws and contributory factors to assist the analysis of indirect controllers; therefore the analysis of indirect controllers is conducted as described in CAST i.e. by identifying the responsibilities related to preventing the crash, unsafe control actions, mental model flaws and context.

Indirect controllers:

Tesla Company:

<u>Responsibilities:</u>

- Responsible to ensure that the vehicles being designed, developed and commercialized are safe and can be safely operated by customers.

Unsafe control actions (UCAs):

- **UCA-5:** Tesla commercialized a version of an SAE 2 automated driving system that can be (mis)used as an SAE 3 automated driving system, and engaged on highway segments with uncontrolled intersections.

Mental model flaws:

- The company believed customers were going to continue monitoring the driving environment while the Autopilot was engaged.
- The company was not aware of the risks associated to the autopilot function and believed that the system was safe.
- The company thought that customers' real driving data was necessary to enhance the automation and therefore they accepted the risk of releasing an early version of the function.

Context:

- The context of their decisions was characterized by the pressure of being a cutting edge technology company that must bring vehicle automation into the market and the legislation and regulatory gaps for vehicle automation.

The NHTSA:

Responsibilities:

- Responsible for the definition and enforcement of vehicle measures and federal vehicle standards to improve vehicle safety.
- Responsible for the definition of guidelines and measures for driver education[17].

Unsafe control actions (UCAs):

- **UCA-6:** The NHTSA did not conduct in-time evaluations of the vehicle standards needs for automated driving system.

[17] The NHTSA has other responsibilities such as conducting research on road safety and establishing accident databases, however for the illustration of the analysis only the two main responsibilities related to this specific crash were considered.

- **UCA-7:** The NTHSA did not establish vehicle regulations on truck side guards which could have been detected by the radar and the camera.
- **UCA-8:** The NHTSA did not define measures to mitigate the human factor issues related to automated driving.

Mental model flaws:

- They thought that new vehicle standards could be defined later, even if regulations were outdated relative to the new vehicle systems already introduced on the roads.
- They were not aware that radars could interpret trucks as overhead road signs and that truck side guards could enhance radar detection of trucks.
- They believed that human drivers were capable of safely operating SAE 2 automated driving systems without additional measures on human driver behavior.

Context:

Finally, the context of their decisions was characterized by the rapid introduction of vehicle automation and the legislation and regulatory gaps for vehicle automation.

5. Generate recommendations

The last step is to use the outputs of the analysis, to generate recommendations that aim at redesigning the system. The following recommendations were created to illustrate the type of recommendations that can be elaborated thanks to the CASCAD analysis.

Recommendations:

- The Tesla Company should review the design process, notably how design assumptions are being made and validated for the vehicle perception strategies (e.g. type of sensors, data fusion choices, false positive avoidance, etc.). They should fix the design features contributing to the crash[18].

[18] The specific design features contributing to the crash were partly addressed by Tesla in the version 8 of the autopilot software (The Tesla Team 2016)

- The Tesla Company should also fix the design of the driver monitoring system[19] to better detect driver's engagement, and redesign the HMI to show the driver what automation perceives.
- The Tesla Company should fix the design of Autopilot to not allow Autopilot's engagement when the vehicle is outside of its operational design limits (e.g. highway intersection).
- The NHTSA needs to evaluate the process in which they review the vehicle standards needs for new technologies and establish new vehicle standards.
- The NHTSA should also review the vehicle measures and federal standards on the other vehicles (e.g. truck side guards) that have the potential to enhance automated driving system's perception and detection.
- The NHTSA should continue studying the driver behavior requirements and human factor challenges introduced by automation in order to propose adapted measures and standards (e.g. driver licenses and education campaigns for automated driving).

5.5 Discussion

This section presents the discussion of the contributions and limitations of three topics: the identified elements specific to road safety, the developed elements to facilitate the application of CAST to vehicle automation, and finally the CASCAD method.

5.5.1 Elements specific to road safety from crash analysis methods

This chapter demonstrated that two elements (the crash description as four phases and contributory factors) from existing crash analysis methods are still relevant for crashes involving automated driving. These elements ensure that some of the specificities of road crashes are incorporated into CASCAD and that practitioners find the notions and language that they expect to see in a crash analysis method.

[19] The version 8 of the autopilot also slows down the vehicle when the driver does not respond to "Hold Steering Wheel" messages within 15 seconds. Further, the driver cannot restart the autopilot system until the car has come to a halt (Plummer 2016)

The crash description as four phases establishes the timeline of a crash independently of the vehicle automation level associated to the vehicles involved in the crash; the description helps the understanding of the crash and organizing the events of the crash for the analysis of the failures and unsafe interactions at the physical level.

The existing driver failure taxonomies were found to include some general categories issued from human information processing models which may be adapted into specific human driver failures for automated driving and specific failures for automation. For example, the failures related to information acquisition, diagnosis prediction, decision-making, and action execution may be appropriate to account for human driver failures during automated driving and for the failures related to the automated driving controller. While it seems feasible to investigate the applicability of human driver failure taxonomies to the human driver failures during automated driving and automation failures, the research in the thesis preferred to shift towards the control flaws classification proposed by Leveson which is suitable for both human controllers and automated controllers. This decision was made because the control flaws classification proposed in STAMP is compatible with CAST and because this classification has already been applied to automated vehicles (Van Eikema Hommes 2012; Abdulkhaleq et al. 2017)

Furthermore, the contributory factors (i.e. the second identified element) included in the HFF framework and DREAM cover factors that will certainly continue to play a role in crashes involving automated driving. Notably for the automated driving SAE levels 1-2 in which the driver is expected to perform a part or all of the dynamic driving task and for the SAE levels 3-4 in which the driver is expected to respond to takeover requests during automated driving. Additionally, in SAE levels 1-4, the driver still has to perform the entire driving task during manual driving.

However, the contributory factors mentioned in the existing crash analysis methods focus on the human driver and on the influence of the vehicle, the infrastructure and the environment components on the human driver. While this makes sense in today's road transport system, vehicle technology is rapidly changing and consequently the knowledge provided by the contributory factors must be extended and updated to include additional contributory factors related to automated driving. Some examples of new contributory factors for the human driver

could be the loss of situation awareness, distraction, overreliance, motion sickness, level of experience with automation, etc. Examples of new contributory factors for automation could be the performance of the vehicle sensors under degraded conditions (e.g. heavy rain, fog, etc.), incorrect information regarding digital maps, and flawed software requirements.

5.5.2 The elements to facilitate the application of CAST on automated driving

Although a CAST analysis could be conducted on crashes involving automated driving without additional guidance, this chapter showed that it is possible to apply STAMP concepts on vehicle automation to develop guidance elements that facilitate the use of CAST on crashes involving automated driving.

The first developed element i.e. the control structure at the physical level, was created to encourage the persons conducting the analysis to explicitly mark the unsafe interactions at the physical level involved in the crash. This structure is the starting point of the analysis and ensures that all the unsafe interactions are identified. The control flaws classification for the human driver controller and the automated controller which was established by examining the control loop of a generic automated driving system's control structure, demonstrated that STAMP can identify control flaws equivalent to those in the human failure taxonomies and additional flaws associated to automated driving, which are not considered in the human driver failure taxonomies. For instance, the human driver flaws related to the inadequate human perception of feedback on another road user and to human missing perception of infrastructure feedback are equivalent to the human driver failures in information acquisition and observation. Furthermore, the automation flaws related to measurement inaccuracies on infrastructure feedback measured by vehicle sensors and sensor inadequate operation represent additional flaws which are not considered in the existing failure taxonomies.

The third developed element was the control structure of the entire road transport system in France which includes the indirect controllers of the system design and operation arranged in six hierarchical levels. The main contribution of this structure is that it extends the scope of the analysis by looking at the high-level controllers that influence the operating process in which crashes take place. The existing methods hint at the influence of higher levels, for instance the

HFF framework considers explanatory elements related to inadequate infrastructure design and DREAM considers organizations, maintenance, vehicle design and road design categories in the organization genotype; however, these methods do not mention the actual high-level controller that influenced the crash being analyzed and the interactions among direct and indirect controllers across all the levels of the transport system. As a consequence, the role of the high-level controllers is not comprehensively examined during the analysis and important unsafe interactions are omitted.

5.5.3 CASCAD

CASCAD indicated that some elements specific to road safety and the guidance elements developed based on STAMP concepts can be integrated into CAST, to extend the CAST method for the analysis of crashes involving automated driving. Furthermore, the application of CASCAD on the Tesla crash, illustrated the relevance and usefulness of the method. The description of crashes as four phases and the control structure at the physical level contributed to the second step of the CASCAD process in which physical failures and unsafe interactions are identified. The analysis of the Tesla driver, the automated controller of the Autopilot function and the truck driver proved that the control flaws classification developed for CASCAD is capable of capturing the causal factors of crashes involving automated driving. While the contributory factors from existing methods were suitable for most of the human driver flaws, they were clearly unfitted for the flaws associated to automation and to human interaction with automation. Therefore, it is recommended that contributory factors associated to automation and to the human interaction with automation should be examined in order to update them.

Furthermore, the analysis of indirect controllers demonstrated that extending the scope of the analysis to explicitly include high-level controllers, allows the identification of additional causal factors associated to the crash. An interesting finding was that the control structure of the road transport system in France partly reflected the control structure of the road transport system in the US (which was the control structure concerned in the Tesla crash analysis) and therefore the control structure of the US transport system was rather simple to build. Nonetheless, the specificities of every road transport system at the moment of the crash need to be taken into account and modeled in order to provide a consistent representation of the system. Finally,

extending the scope of the analysis and generating recommendations that also address the higher levels of the system, requires sense of change and commitment across all the levels of the system (the government, insurance companies, vehicle manufacturers, infrastructure companies, the persons conducting crash analyses, etc.). Some of the actions that can support this view include: raising awareness of the importance of considering the entre sociotechnical system in crash analyses and prevention, incorporating causal factors that address aspects beyond the operating process into the vocabulary used by the road safety community, and creating variables related to the high-level causal factors within the accident databases.

The main limitation of the CASCAD method is the lack of data on real crashes involving automated driving systems to validate the CASCAD method. The Tesla crash illustrated the application of the method, however, more crashes and the access to all the information required for the analysis, are needed to validate the application of the method. Additionally, comparisons with the existing crash analysis methods are also necessary in order to demonstrate that CASCAD is more suitable for crashes involving automated driving. Lastly, the practitioner's opinion on CASCAD also has to be examined; their thoughts on the usefulness of the CASCAD method and potential improvements can be integrated to enhance CASCAD and to provide a method that meets practitioners' needs.

5.6 Conclusions

The crashes involving automated driving will most likely require new methods based on systems theory to facilitate their analysis. An accident analysis method celled CAST has been identified as a potential candidate, however, the lack of industry-specific guidance may prevent CAST from being adopted by road safety practitioners. This study creates an accident analysis method for crashes involving automated driving called CASCAD, which incorporates road safety-specific elements from traditional crash analysis methods and elements to facilitate the analysis of automated driving systems, into CAST.

The findings of this study show that some elements from traditional crash analysis methods are still relevant for the analysis of automated driving. Moreover, STAMP can be applied on an

automated driving system in order to generate usage guidance elements for road safety practitioners. These elements are able to coexist with CAST in the CASCAD method.

The methodology proposed in CASCAD was illustrated using available data from the Tesla crash in May 2016. Although, the illustration does not intend to substitute the complete analysis of the crash, it demonstrated that CASCAD is useful for the analysis of automated and human controllers, as well as for the analysis of the high-level controllers of the road transport sociotechnical system.

The development of more guidance elements is recommended, especially for the contributory factors related to the human behavior in automated driving and to the factors that influence vehicle automation. Furthermore, the application of CASCAD on crash investigations involving automated driving and the comparison with the outputs of traditional methods, must be performed in order to validate whether or not CASCAD assists a more complete understanding. Lastly, road safety practitioners should be consulted to identify if CASCAD meets their needs and potential improvements.

5.6.1 Future work

Two future research opportunities were identified for this chapter:

- Improve CASCAD: The identified elements from existing crash analysis methods and the newly developed guidance elements could be enhanced to further improve CASCAD. The empirical evidence on vehicle automation from the literature should be monitored to update the list of contributory factors related to crashes involving vehicle automation. The results of large-scale field operational trials such as the Drive Me project in which 100 vehicles equipped with automated driving systems drive on open roads (“Drive Me” 2017) and the L3Pilot project, will provide insights on the new interactions and causal factors brought by automated driving. Also, the control flaws classification could be reviewed by a group of experts on human factors and vehicle systems, to verify and complete the categories of the classification. The control structures of the road transport system in other regions could be built to model the specificities of other the road transport systems. Lastly,

CASCAD could be introduced to the practitioners of road crash analyses in order to get their opinions on how to improve CASCAD.

- Validate CASCAD: Real crashes involving automated driving will provide the necessary data to apply and validate CASCAD. Cooperation with the organizations in charge of in-depth crash analyses could be established in order to have access to all the necessary information for the CASCAD validation. Additionally, the outputs of CASCAD analyses and the outputs of analyses based on existing methods could be compared to further validate CASCAD. Finally, practitioners could also apply CASCAD and provide information to validate that CASCAD offers adequate guidance and assistance for the analysis of crashes involving automated driving.

Résumé chapitre 6: Discussion

Le chapitre 6 constitue la discussion des résultats généraux de la thèse, les résultats individuels se retrouvant déjà dans la partie discussion des chapitres 3 à 5. Après avoir présenté un rappel des résultats des chapitres précédents, les quatre principales contributions de la thèse sont décrites :

- Apports des méthodes fondées sur STAMP aux trois questions de recherche.
- Représentation du système de transport routier comme une structure de contrôle.
- Extension du périmètre de l'analyse et identification d'un ensemble plus vaste des facteurs contribuant aux accidents.
- Modifications développées pour l'application des méthodes STPA et CAST au véhicule autonome et à la sécurité routière.

Enfin, ce chapitre détaille les considérations méthodologiques de la thèse par rapport aux structures de contrôle développées dans les chapitres, à la validité des résultats et à leur généralisation possible (pertinence des résultats pour d'autres systèmes d'automatisation, pour la sécurité routière hors véhicule autonome et pour d'autres questions sécuritaires tel que la conception).

Chapter 6: Discussion

6.1 Chapter overview

This chapter begins with a summary of the research findings presented in the thesis. Then, the four main contributions of the research as a whole[20] (i.e. the application of a STAMP, STPA and CAST approach to vehicle automation and road safety) are described:

1. The implications of STAMP-based methods for the three research questions.
2. The representation of the road transport system as a control structure.
3. The larger scope of the analysis and identified causal factors.
4. The modifications developed to apply STPA and CAST on vehicle automation and road safety.

Lastly, the methodological considerations of the research regarding control structures, the validity of results and the generalization of the findings, are provided.

6.2 Summary of findings

The first chapter introduced the context of the thesis which involves the development and introduction of vehicle automation into the road transport and its implications on road safety. Three research questions regarding the safety benefit assessment, trial safety and crash accident analysis of automated driving systems, were identified from several industry challenges across the phases of vehicle's development and operation. The need for a suitable conceptual framework to address the research questions and the literature evidence that points out towards systems theory as the next conceptual paradigm shift were underlined. Finally, the aims and approach of the thesis were stated.

Chapter 2 presented an overview of the literature on vehicle automation, road safety and the three most popular systems theoretic approaches. Furthermore, the system theoretic approach

[20]The individual contributions of the research are presented in the discussion section of chapters 3-5.

selected as the conceptual framework for the thesis i.e. the model STAMP and its associated methods STPA and CAST, was described in detail.

A contribution to the safety benefit assessment of automated driving systems was provided in chapter 3. As seen in figure 46, the contribution consisted of the estimation of a highway pilot system's target population and a set of questions derived from safety requirements identified through an STPA analysis, which aim at assisting the evaluation of assumptions related to the assessment of the effects of the highway pilot system and infrastructure on safety (the direct safety mechanisms 1 and 2 among the 9 mechanisms). The results of this chapter indicated that the target population addressed by the highway pilot system in France is around 4,6% of all crashes, 3,8% of road fatalities, 3,3% of road users injured and hospitalized and 3,6% of road users injured and not hospitalized. Moreover, results showed that STPA can be used to model automated driving systems, to analyze the human driver and the automated controllers, and to identify safety requirements and refined safety requirements on system's behavior (notably requirements on the feedback and process models). The questions defined based on the safety requirements and refined safety requirements aim to further assist the evaluation of safety mechanisms 1-2 by targeting the unsafe interactions across all the phases of system operation.

Chapter 4 provided a framework to ensure the safety of trials involving automated driving systems. As observed in figure 46, the framework was established by structuring the safety requirements from a first STPA analysis on the vehicle trial process and from a second STPA analysis on a vehicle trial involving a highway pilot system, into five sections:

- Section 1: Definition of policies and resources for the development of vehicle technology and vehicle trials.
- Section 2: Establishing orientations for vehicle technology development and vehicle trials.
- Section 3: Approval of vehicle trials.
- Section 4: Design and development of vehicle trials.
- Sub-section 4.1: Trial organization and preparation.
- Sub-section 4.2: Trial data.
- Subsection 4.3: Safety and compliance of the trial.

- Section 5: Vehicle trial operation.
- Sub-section 5.1: Safety related to the maturity of the vehicle technology being tested.
- Sub-section 5.2: Safety related to the vehicle trial operation.
- Sub-section 5.3: Trial operation data.

The findings of chapter 4 further illustrated how STAMP and STPA can be used to model both the operational and the higher levels of the vehicle trial system, extend the scope of the analyses and identify safety requirements that address a larger set of causal factors.

Chapter 5 introduced a method for the analysis of crash accident involving automated driving, called CASCAD. As observed in figure 46, CASCAD extended CAST by incorporating elements from existing crash analysis methods and newly developed guidance elements to assist the analysis of automated driving. On the one hand, this chapter showed that some of the elements from existing crash analysis are still relevant for crashes involving automated driving systems; however, other elements such as the human failure taxonomies are not completely adapted to capture flaws related to the new interactions brought by automation. On the other hand, it was demonstrated that STAMP concepts can be applied to develop three guidance elements: the control structure of crashes at a physical level, the control flaws classification for the human driver and automation, and the control structure of the entire road transport system, which intend to facilitate CAST analyses of crashes involving automated driving. Lastly, the CASCAD application[21] on the Tesla crash illustrated the usefulness of the method to facilitate the use of CAST, to capture causal factors related to automated driving, and to extend the scope of the analysis by including the higher-level controllers of the road transport system.

[21] The application of CASCAD on the Tesla crash does not intend to substitute a complete crash analysis. The analysis was based on the information on the internet and the NHTSA official reports; thus important information was missing. The aim of this application was to serve as an example to illustrate the method.

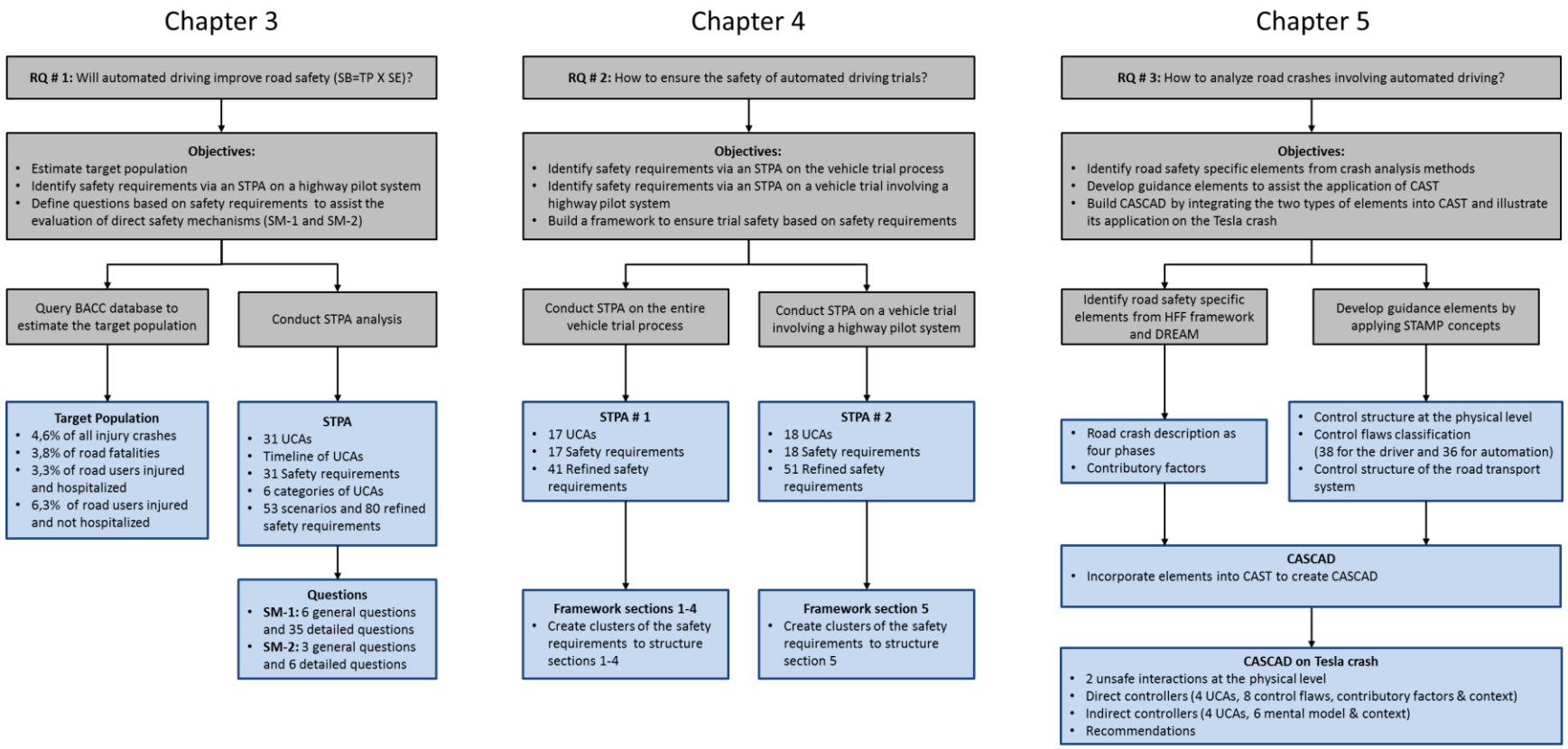


Figure 46 – Summary of findings in chapters 3-5

6.3 Contributions

6.3.1 The implications of STAMP-based methods for the three research questions

The separate findings regarding the implications of STAMP-based questions for the road safety benefit assessment, trial safety and accident analysis of automated driving are detailed in the discussion sections of chapters 3-5. Nevertheless, the main implications of these findings are stated below:

- The process presented in chapter 3 showed that the results of STPA analyses allow the elaboration of questions that facilitate the evaluation of the direct safety mechanisms (1-2) for a highway pilot system. The elaborated questions will be used in upcoming field operational trials (notably the field operational trials in the L3Pilot project) to facilitate the safety benefit assessment of automated driving systems.
- The framework provided in chapter 4 illustrated how STPA analyses help ensuring the safety of automated driving trials by examining not only the low levels of the system but also higher levels such as the government, funding agencies, and the multiple actors within vehicle companies. The framework will be applied to ensure the safety of future vehicle trials at Renault.
- The CASCAD method developed in chapter 5 showed that CAST can be extended to better support the analysis of crashes involving automated driving. CASCAD will help automakers (and other stakeholders of the road transport system) conduct retrospective evaluations of automated driving system by providing an accident analysis method adapted to the crashes involving automated driving.

A key implication of the thesis findings is that the STAMP-based methods offer a new way to model and analyze automated driving systems, which can help the automotive industry and the other stakeholder concerned by road safety, to cope with the complexity of the road transport system and the new changes and interactions introduced by automation.

6.3.2 Modeling the road transport system as a control structure

To address the three research questions that motivated the research of the thesis, the system considered in each question had to be modeled before being analyzed. Therefore, the first

contribution of the thesis is the application of STAMP concepts to model the human, technical and organizational factors of the road transport system as a hierarchical control structure comprising multiple levels.

Chapters 3-5 demonstrated how the STAMP concepts can be used to model the human driver, automation, the vehicle, the environment, and their interactions in the system operation level as a control structure. The notions of control actions and feedback loops were very useful to represent the process in which the human driver controller and the automated controller receive feedback from the controlled process and provide control actions on the vehicle to enforce safety constrains and remain a safe behavior. Further, the control actions and feedback between these two controllers (e.g. human ADS engagement validation and automation's takeover request notifications) also captured their interactions. Explicitly pointing out and illustrating the interactions of a given system facilitated the evaluation of the control structure's comprehensiveness and the subsequent understanding of the system. For instance, when the control structure of the highway pilot system was shown to the system's designers, they could easily notice whether or not the intended interactions were represented. This enabled designers to correct and complete the structure and to become aware of previously unnoticed interactions. Moreover, chapter 4 also modeled the trial supervisor and the trial staff participating in trial operations.

Chapters 4 and 5 illustrated how the STAMP concepts can be used to model the higher level controllers (i.e. stakeholders) of the road transport system. For example, the entire control structure for the vehicle trial process built in chapter 4 encompassed the government, funding agencies, the multiple controllers within the vehicle manufacturer and the controllers at the vehicle trial operating process. Additionally, the road transport control structure established in chapter 5 includes international stakeholders, associations, road infrastructure companies, automakers, driving schools etc.

The contribution of modeling the entire road transport system as a control structure can be observed by looking at the conceptualization of the Safe System approach presented in section 2.3.4. As observed in figure 6, the stakeholders such as legislators, corporations, designers, enforcers, users, etc. and the components of the operating process (e.g. road users, speed,

roads, and post-crash response) are considered in the representation of the Safe System approach. However, compared to hierarchical control structures, the specific interactions of the stakeholders and the components of the operating process are not clearly and explicitly detailed in figure 6.

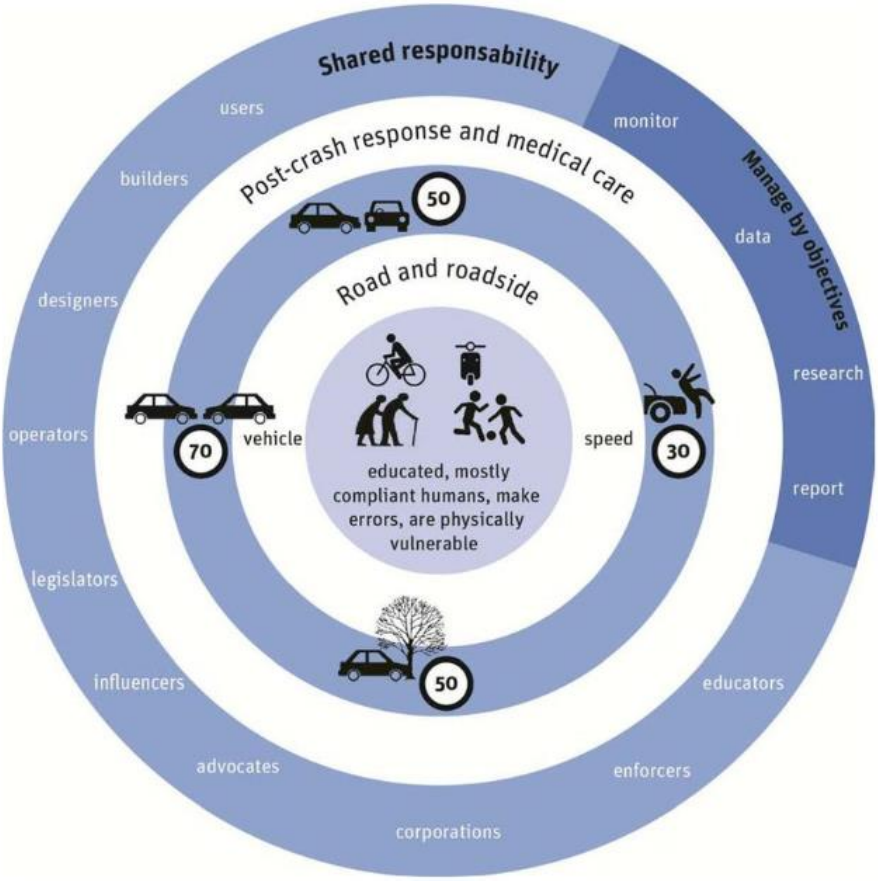


Figure 6 - Conceptualization of the Safe System (ITF 2016)

Finally, an important advantage of the control structures built in the thesis is that they overcome the challenges of modeling the microscopic perspective of the operating process and the macroscopic perspective of the higher levels of the entire road transport system, by using a unique representation of controllers at all levels. The use of a common representation for all the controllers enabled the coexistence of the human, technical and organizational factors encompassed in the whole system and established relationships between the microscopic and macroscopic perspectives. Furthermore, a common representation of all the system levels can potentially facilitate the communication between people from various scientific disciplines like

engineering, human factors, political sciences, etc. who work at different levels of the structure. For instance, the engineers developing the technical aspects of the automated driving system and the human factors experts designing the HMI interfaces, can use the common representation to work together on the whole process in which the vehicle sensors measure and send feedback to automation, next automation sends feedback to the HMI, which in turn displays feedback to the human driver.

6.3.3 The modifications developed for the application of the systems theoretic approach

The third main contribution was the modifications developed to facilitate and to adapt the application of the systems theoretic approach comprising STAMP, STPA and CAST, throughout the thesis. These modifications can be divided into two types:

1. The modifications created to enhance the application of the methods and the way results are reported (or displayed).
2. The modifications developed to adapt the approach to the road safety domain.

In chapter 3, several modifications were developed in order to optimize the processing time of the analysis and to display the STPA results in a comprehensible way to people unfamiliar with STPA. The 32 unsafe control actions defined for the human driver and the automated controller were grouped into six categories of unsafe control actions to reduce the number of inputs for the STPA step 2. This showed that it is possible to classify similar unsafe control actions to optimize the processing time of the analysis. Additionally, the timeline that graphically displayed the distribution of control actions and unsafe control actions across the multiple phases of the automated driving system's operation (illustrated in figure 27) facilitated discussions over the STPA step 1 results with people that did not conduct the analysis. Regarding STPA step 2, the use of high-level control flaws classes and color-coding (see appendix A) helped to make results understandable to inexperienced STPA users.

In chapter 5, the CAST method was modified and extended to create CASCAD, an accident method more adapted to the road safety domain. This proved that the "generic" nature of STAMP-based methods can be adapted to better meet the needs of a specific industry. To build

CASCAD, some of the elements from current road crash accident analysis methods were incorporated into CAST, which showed that some features of existing industry-specific methods can be compatible with CAST. Lastly, the STAMP-based analysis used to develop guidance elements for the application of CAST to crashes involving automated driving systems, illustrated how STAMP concepts can be employed to create industry-specific elements.

6.3.4 Extending the scope and findings of analyses on road safety

Once the representation of the system has been established, the next step is to analyze the system. Accordingly, the second main contribution of the thesis is related to the larger scope and potentially[22] larger set of findings provided by the STAMP-based, STPA and CAST analyses. Traditionally, the main focus of road safety analysis has been on the driver-vehicle-environment system (see section 2.3.3) and therefore most of the identified contributory factors are associated to those components at the microscopic level e.g. driving under the influence, distraction, tire blowout, reduced friction, etc. A consequence of explicitly representing the entire sociotechnical system is that it naturally extends the scope of the analysis; when the higher level controllers and the influences of their decisions and actions become visible, the analyst is encouraged to look beyond the operating process comprising the driver, vehicle and environment components. Additionally, the results obtained in the thesis showed the large set of causal factors that can be identified with STAMP-based, STPA and CAST analyses across the entire control structure.

At the microscopic level, the analyses identified causal factors which are already known thanks to today's methods such as failure of electromechanical components, unsafe human driver behavior, incorrect feedback provided by vehicle sensors, inadequate actuator operation, delays, etc. Further, the systems-theoretic approach identified new causal factors related to automation which are not comprehensively addressed by current methods; new causal factors included inadequate feedback, software errors, flawed software requirements, design errors and unsafe interactions. For instance, in the Tesla crash (described in chapter 5), the camera

[22] It seems evident that a systematic analysis with a larger scope will potentially lead to a larger set of findings; nevertheless, this assumption still has to be verified by comparing STAMP-based analyses with other analyses.

provided incorrect feedback indicating that there was no obstacle because the camera was unable to distinguish the white trailed of the truck against a clear blue sky. On the other hand, the radar detected an obstacle but provided inadequate feedback indicating that the trailer was an overhead traffic sign. In turn, the inadequate feedback was caused by flawed software requirements defined by the software developers to tune out signals and avoid false braking events. Finally, a design error was involved when the designers of the automated driving system allowed the operation of the autopilot feature on uncontrolled highway intersections and relied on the driver for safety.

Furthermore, the safety requirements identified in chapters 4 and 5 demonstrated that the influences of the macroscopic levels of the system on the operating process can be explicitly stated to establish links between causal factors at the microscopic levels and the higher-levels of the system. For example, the causal factors related to the human driver's misuse of automated driving systems are influenced by the automaker's design and validation of the system and the vehicle regulations defined by the government. Additionally, the findings indicated that most of the identified causal factors related to high level controllers encompass inadequate or missing feedback provided by lower levels and inconsistent mental models of vehicle technology and safety problems. For example, the government may establish inadequate regulations for automated driving trials because they are unaware of the risks associated to such trials. In turn, their inadequate representation of the risks associated to automated driving trials is partly due to the feedback that they receive from the lower levels of the system. Nonetheless, empirical data on the interactions and processes across the higher levels of the road transport system are needed to confirm that the identified causal factors are real and to further examine the reasons behind them. For instance, interviews should be conducted to study and compare the governments and automakers' representation of vehicle trial safety and the feedback channels and communication between the two levels.

6.4 Methodological considerations

6.4.1 Control structures

The findings of the thesis showed that it is necessary to describe the system and its control actions with an appropriate level of detail for the analysis being performed. The first attempts in building control structures at the microscopic level incorporated detailed descriptions of the control actions provided by the human driver and automation. For instance, the driver provides acceleration, braking and steering to control the motion of the vehicle. This decomposition of the control actions for vehicle motion was too refined for the analyses regarding the safety benefit assessment, trial safety and crash analysis; it increased the number of unsafe control actions and resulted in safety requirements that included too many details which were not particularly relevant for the analyses. For example, the driver's actions to control the motion of the vehicle were initially divided into the driver's acceleration, braking and steering actions, and generated safety requirements such as:

- The driver must not provide acceleration when the safety distance to a vehicle in front is violated.
- The driver must provide braking when the safety distance to a vehicle in front is violated.

In the subsequent control structures, control actions were described with less detail, for example, the driver's acceleration, braking and steering actions were grouped into: the driver provides control of the vehicle. As a result, the number of unsafe control actions decreased and the resulting safety requirements included an appropriate level of description (e.g. the driver must provide adequate control of the vehicle). Therefore, it is recommended to always start with high level descriptions of the system and the control actions, and add the details later.

Additionally, the initial control structures at the microscopic level showed that the distinction between control actions and feedback is not always clear. For example, the interaction in which automation sends takeover request notifications to the human driver was first considered as a feedback loop because it was represented as an upward arrow in the control structure. Accordingly, this feedback loop was not examined in the STPA step 1 (which analyzes control

actions) but in the STPA step 2. However, as the scenarios were being generated in the STPA step 2, it was observed that the takeover requests sent by automation had a great influence on the driver's takeover validation response and thus takeover request notification could also be considered as a control action. Ultimately, it was determined that seeing the takeover requests as a control action instead of a feedback loop was more suitable for the analyses because it generated more emphasis on the takeover request. While the consequences of classifying an interaction as a control action or a feedback loop are not serious as long as the interaction is comprehensively examined (in the STPA step 1 or STPA step 2), it is interesting to note that some interactions can sometimes be classified as both. This issue should be discussed with the different members performing the STPA in order to select the category which is more suitable for the analysis being performed.

Leveson clarifies that control is a very broad notion in STAMP; there can be physical, human, organization and social controls. The control structures established in this thesis confirm the findings of (Salmon, Read, and Stevens 2016), who concluded that the notion of control is not straightforward; in fact, as the levels of the road transport control structure increase, it becomes harder to grasp the control mechanisms and their influences on the system. While the control mechanisms and effects at the low operation level such as the roads' physical constraints on the road users and the drivers' (or automation's) actions to control the motion of the vehicle, are easily observed; the high-level control mechanisms and their effects such as the regulations and standards on vehicle design, are not obvious. The large view of control and the different degreed of influence provided across the entire road transport system should be explained to inexperienced STAMP users. As suggested by (Salmon, Read, and Stevens 2016), a distinction can be made between direct controls at the lower levels and the influencing mechanisms at the higher levels, in order to help novice users see the broad range of control mechanisms and the links between them.

6.4.2 Validity of results

Although documentation was available, and experts, designers and other employees contributed by providing additional information on the systems and by validating the control structures, the analyses were mainly performed by one researcher who was not an expert on

automated driving systems. Therefore, the validity and reliability of the findings might be limited by the degree of expertise and knowledge of the researcher; it is possible that a group of experts will come up with a larger set of scenarios and safety requirements. This limitation was partially addressed by having an expert on the system verify the findings of the STPA and STAMP-based analyses. Moreover, the plausibility of numerous elaborated scenarios was also verified with designers and experts on the system. That being said, the intention of this thesis was not to provide a set of exhaustive and valid results but mainly to find a suitable conceptual framework and methodological approach to examine the influence of vehicle automation on road safety.

The ideal process to further verify the validity of the research findings would have been to compare the findings with results obtained through traditional methods (e.g. FMEA or HAZOP with STPA and HFF or DREAM with CAST). Unfortunately, the FMEA analysis on the highway pilot system was not available before the end of the research period and there were no HFF or DREAM analyses on crashes involving automated driving, therefore it was not possible to make the comparison. Additionally, it is worth mentioning that to make a fair comparison, the scopes of the analyses should be the same. Nonetheless, the scope of the STPA analyses conducted in the thesis is larger than the scope of FMEA analyses normally performed in the vehicle industry, which raises the question whether or not a FMEA is suitable to perform an analysis on an entire sociotechnical system. Moreover, the HFF and DREAM analyses include causal factors related to the higher levels of the sociotechnical system such as inadequate infrastructure design or work pressure, but they do not set the scope of the analysis to explicitly depict the whole sociotechnical transport system.

6.4.3 Generalization of the findings

Given that the control structures, scenarios and safety requirements have a broad and high-level description, many of the thesis results are applicable to other automated driving systems. For example, the safety requirement in which the automated controller must not provide control of the vehicle during manual driving, is valid for all automated driving systems. Another example is the safety requirement in which the driver must not validate the takeover request and put the vehicle in an unsafe situation. Nevertheless, the safety requirements for the

highway pilot system are not comprehensive for all the automated driving systems; an STPA analysis on another system will most likely identify a few additional safety requirements—and maybe exclude some of the highway pilot safety requirements—which reflect the specific characteristics of the given system. For this reason, it is suggested that the specificities of every automated driving system and the conditions of a given trial, should be considered and incorporated into the analysis. Accordingly, the approach used on the thesis to apply STAMP and STPA to automated driving systems can be used on other automated driving systems.

Furthermore, the results involving the higher levels of the road transport sociotechnical system, are valid not only for automated driving systems but for all vehicle systems. This is illustrated by some examples of the identified safety requirement on trial safety (chapter 4). For instance, the safety requirements related to the government's proper representation and understanding of the vehicle technology being tested and the safety of the vehicle trial; moreover, the safety requirements associated to the feedback on vehicle technology and on vehicle trial conditions that the automaker must provide to the government. These two safety requirements are relevant to all vehicle systems.

Finally, the objectives of the thesis deliberately excluded the design phase, in order to focus on the validation, deployment and retrospective evaluation phases; and therefore the safety requirements and scenarios defined in chapters 3-5, were not employed to change the design of the system. However, these results could be used to improve system design (including the automated driving system and the entire transport system) in order to make it safer. Further, other STAMP-based methods which were not employed in the thesis could be applied to other safety issues regarding vehicle automation such as using STPA-sec for cyber security, using STECA for early design (Fleming 2015) and leading indicators to monitor the state of the system (Leveson 2015).

Résumé chapitre 7 : Conclusions et perspectives

Ce dernier chapitre conclue les travaux menés dans la thèse et présente les perspectives de recherches. Au regard de l'objectif initial de la thèse, les résultats obtenus sur les trois questions de recherche permettent de conclure que l'approche STAMP est adaptée à l'étude des implications du véhicule autonome sur la sécurité routière. Pour aller plus loin, des perspectives de recherche sont données afin de continuer le travail sur les questions de recherche et d'étudier les nouveaux types de systèmes et les nouvelles questions sécuritaires en rapport avec le véhicule autonome. Plusieurs suggestions sont également faites pour diffuser et encourager l'adoption de l'approche au sein de la communauté de la sécurité routière.

Chapter 7: Conclusions and future work

7.1 Conclusions

The research conducted in this thesis aimed to examine the safety benefit, trial safety and the accident analysis of automated driving, by applying a systems theoretic approach (i.e. the STAMP model and associated methods STPA and CAST). The STAMP-based approach was selected to independently address the three issues by modeling and analyzing the multiple levels of the entire road transport system and their interactions.

Regarding the safety benefit assessment, the estimates of the crash target population addressed by the highway pilot system were 4,6% of all crashes, 3,8% of road fatalities, 3,3% of road users injured and hospitalized and 3,6% of road users injured and not hospitalized. Moreover, the questions derived from the identified safety requirements offer structured and comprehensive assistance for the evaluation of direct safety mechanisms 1-2. The upcoming real-driving field operational trials in the EU-funded L3pilot project will provide an opportunity to investigate the usefulness of these questions for the evaluation of the direct safety mechanisms and quantifying the safety benefits expected for some automated driving functions.

Furthermore, the framework based on the safety requirements identified through two STPA analyses on the vehicle trial process and on a trial operation involving a highway pilot system provides a scheme to ensure trial safety beyond the operating process; it covers safety across the multiple levels of the vehicle trial system e.g. the government, funding agencies, the actors within vehicle manufacturers, and trial operation. The framework will be applied by Renault on future vehicle trials necessary for the deployment of automated driving systems.

The CASCAD method extended CAST by integrating elements specific to the road safety domain and newly developed elements to facilitate the analysis of crashes involving automated driving. The application of CASCAD on the Tesla crash illustrated the capability of the method to capture causal factors related to automation for both the human controller and the automated controller. Additionally, it also led the analysis to consider the indirect controllers at the higher

levels of the road transport system such as the NHTSA and the Tesla Company. However, the application of CASCAD on crashes in which all the accident investigation information is available, and the comparison of CASCAD analyses with analyses based on traditional methods (e.g. the HFF framework and DREAM), are still needed to further explore CASCAD's suitability and advantages for the retrospective evaluation of automated driving systems.

Based on these findings, it is concluded that the application of a STAMP-based approach on the safety benefit assessment, trial safety and accident analysis, showed that STAMP, STPA and CAST provide a suitable conceptual framework for the analysis of the automated driving system's implications on road safety. It is worth noting that the starting assumption off the thesis was to choose STAMP over the other systems theoretic approaches (i.e. the Risk Management Framework and FRAM described in chapter 2) and therefore the other approaches may also provide an appropriate conceptual framework for automated driving and road safety. However, the comparison between the three approaches and a critical analysis of their contributions and limitations was beyond the objectives of this thesis.

7.2 Future work

7.2.1 Progression from thesis

This section describes the possibilities for furthering the research on the application of a STAMP-based approach to the safety benefit assessment, trial safety and accident analysis of automated driving systems.

7.2.1.1 Progression from safety benefit assessment

A key limitation of chapter 3 was that the questions derived from safety requirements have not been examined using empirical data; the relevance of these questions needs to be explored by applying the questions on real studies in which the interactions among the human driver, automation and the driving environment can be observed. The upcoming field operational trials of the L3Pilot project in which the highway pilot system considered in this thesis will be operated on open roads, offer an opportunity to observe these interactions and to assess the expected benefits of the system. The use of this data to tackle the questions should facilitate the evaluation and quantification of direct safety mechanisms 1-2. Moreover, the application of

STPA to derive question relative to the other safety mechanisms (3-5) encompassed in the risk dimension, should also be explored. For instance, the interactions of the control structures could be modified to include the interactions with other road users in terms of communication and to derive questions on the modification of interactions between road users (safety mechanism 5). Finally, future research should also consider the integration of the STPA method into the overall evaluation of the 9 safety mechanisms and the other two road safety dimensions i.e. exposure and accident consequences.

7.2.1.2 Progression from automated driving trial safety

The framework established in chapter 4 provides guidelines to ensure the safety (across all system levels) of vehicle trials necessary for the deployment of automated driving systems. The next natural step is to use the framework during vehicle trial operations at Renault in order to contribute to the safety of such trials. Furthermore, the specificities of different trial conditions and the prototypes being tested should be incorporated into the framework.

7.2.1.3 Progression from analysis of crashes involving automated driving

Although the illustration of CASCAD using the available information on the Tesla crash showed the potential of the method to capture causal factors related to automated driving, CASCAD should be applied on more crashes on which all the necessary information to conduct the analysis is available. Moreover, the CASCAD analyses could be compared with the analyses based on existing crash analysis methods to examine the suitability and advantages of CASCAD for the retrospective evaluation of automated driving systems relative to the existing methods. Lastly, CASCAD could be introduced to practitioners such as the persons that conduct in-depth crash analyses, in order to evaluate the method's perceived usefulness, resource demands, understandability, flexibility, etc. Also, practitioners could provide recommendations to improve the method and to further adapt it to the road safety domain.

7.2.2 Examine new automated driving systems and new applications

Whilst the findings of the thesis demonstrated the application of a STAMP-based approach for a highway pilot system (SAE 3) and an autopilot function (SAE 2); there is clearly a potential to apply the approach on other automated driving systems and other driving environments. The

application of the STAMP-based approach for the safety benefit assessment, trial safety and accident analysis of the other automated driving systems being developed at Renault and "analogous" systems being developed by other automakers, should be examined. On the other hand, new applications of the STAMP-based approach on safety-related issues could also be explored. For example the use of STPA and STECA (a STAMP-based method for the early concept analysis) on the design of automated driving systems and the use of STPA-sec (a STAMP-based analysis for cyber-security) on cyber security.

7.2.3 Encourage the adoption of a STAMP-based approach for road safety

This thesis showed the potential of a STAMP-based approach to provide a suitable conceptual framework for automated driving and road safety. The next step is to disseminate the approach and encourage the road safety community to adopt it. To this end, the following suggestions were identified:

- Provide data on STAMP-based applications: Evidence on the usefulness and benefits of using a STAMP-based approach is needed in order to persuade the road safety community to invest in STAMP. Applications using a STAMP-based approach on both automated and traditional driving and comparisons with applications based on current methods could provide data that demonstrate the value of the new conceptual framework. The data regarding the STAMP-based methods' learning curves of the analysts, the resources spent in the analyses (e.g. number of people, time) could also be documented to indicate the costs of shifting to a new approach.
- Increase the visibility of STAMP-based approaches in road safety circles: The members of the road safety community need to be aware that STAMP-based approaches exist and that they are being applied to road safety. The advocates of systems theory perspectives to road trauma should use their platforms (journal papers, conferences, networks, etc.) to make the STAMP-based approaches visible in road safety circles and to raise awareness about the new methods. Integrating STAMP-based methods into standards and official procedures could also help to increase visibility; while efforts are being made to include STPA in the standard ISO 26262 for the functional safety of automotive

electric/electronic systems (Suo et al. 2017; Abdulkhaleq et al. 2017), the integration of the methods at the macroscopic level also needs to be addressed.

- Bring the road safety and STAMP communities together: Another way to encourage adoption is to get the road safety and STAMP communities to talk and to work on common issues. Although the automotive industry is vastly involved with the STAMP community (they are very vocal about their use of STPA and active participants in STAMP workshops), the other members of the road safety community are not as eagerly implicated in the activities organized by the STAMP community. As a starting point, the supporters of the Safe System approach could be reached to discuss the contributions that STAMP-based methods may bring to their initiatives and possible cooperation.
- Develop industry-specific extensions: The last suggestion that has been identified is to develop industry-specific extensions of the STAMP-based methods adapted to the needs of road safety. This issue has partly been addressed in chapter 4 with the development of CASCAD. Nonetheless, more extensions should be established, notably for the STPA method, taxonomies of contributory factors and data collection. Additionally, the efforts to adapt the methods have to be conducted with the participation of practitioners and researchers from the road safety domain in order to ensure that the extensions are shaped to their needs and perceived as useful.

References

Abdulkhaleq, Asim, Stefan Wagner, Daniel Lammering, Hagen Boehmert, and Pierre Blueher. 2017. “Using STPA in Compliance with ISO 26262 for Developing a Safe Architecture for Fully Automated Vehicles.” *arXiv:1703.03657 [Cs]*, March. http://arxiv.org/abs/1703.03657.

Alessandrini, Adriano, Andrea Campagna, Paolo Delle Site, Francesco Filippi, and Luca Persia. 2015. “Automated Vehicles and the Rethinking of Mobility and Cities.” *Transportation Research Procedia* 5: 145–60. doi:10.1016/j.trpro.2015.01.002.

Assailly, J.P. 1993. *Les Jeunes et Le risque.Une Approche Psychologique de L’accident,*. Vigot. Paris.

Bainbridge, Lisanne. 1983. “Ironies of Automation.” *Automatica (Journal of IFAC)* 19 (6): 775–79. doi:doi:10.1016/0005-1098(83)90046-8.

Bird, Frank E., and George L. Germain. 1996. *Practical Loss Control Leadership*. Rev. ed. Loganville, Ga: Det Norske Veritas (U.S.A.).

Brown, Austin. 2013. “Automated Vehicles Have a Wide Range of Possible Energy Impacts.” presented at the TRB 2nd Annual Workshop on Road Vehicle Automation, July 16.

Carter, A., A. Burgett, G. Srinivasan, and R. Ranganathan. 2009. “Safety Impact Methodology (SIM): Evaluation of Pre-Production Systems.” In . Stuttgart, Germany.

Cunningham, M., and M.A. Regan. 2015. “Autonomous Vehicles: Human Factor Issues and Future Research.” In . Gold Coast, Queensland.

“DAVI – Dutch Automated Vehicle Initiative.” 2017. Accessed May 18. http://davi.connekt.nl/.

Draskóczy, M., O.M.J Carsten, and R. Kulmala. 1998. “Road Safety Guidelines.” Delverable B5.2. CODE project, Telematics Application Programme.

“Drive Me.” 2017. Accessed May 18. http://www.volvocars.com/intl/about/our-innovation-brands/intellisafe/autonomous-driving/drive-me.

Elvik, Rune, ed. 2009. *Handbook of Road Safety Measures*. 2., [Rev.] ed. Bingley: Emerald.

———. 2013. “Risk of Road Accident Associated with the Use of Drugs: A Systematic Review and Meta-Analysis of Evidence from Epidemiological Studies.” *Accident Analysis & Prevention* 60 (November): 254–67. doi:10.1016/j.aap.2012.06.017.

Endsley, Mica R. 1995. “Toward a Theory of Situation Awareness in Dynamic Systems.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37 (1): 32–64. doi:10.1518/001872095779049543.

Eskandarian, Azim, ed. 2012. *Handbook of Intelligent Vehicles*. Springer Reference. London ; New York: Springer.

European Commission, and DG for Mobility and Transport. 2015. *Road Safety in the European Union: Trends, Statistics and Main Challenges, March 2015*. Luxembourg: EUR-OP.

Evans, Leonard. 2004. *Traffic Safety*. Bloomfield, Mich: Science Serving Society.

Fleming, Cody Harrison. 2015. "Safety-Driven Early Concept Analysis and Development." Massachusetts Institute of Technology (MIT).

Gasser, Tom M., and Daniel Westhoff. 2012. "BASt-Study: Definitions of Automation and Legal Issues in Germany." presented at the Road Vehicle Automation Workshop.

Habib, Karim. 2017. "ODI RESUME." PE-16007. NHTSA.

Hagenzieker, Marjan P., Jacques J.F. Commandeur, and Frits D. Bijleveld. 2014. "The History of Road Safety Research: A Quantitative Approach." *Transportation Research Part F: Traffic Psychology and Behaviour* 25 (July): 150–62. doi:10.1016/j.trf.2013.10.004.

Hakkert, A.S., V. Gitelman, and M.A. Vis. 2007. "Road Safety Performance Indicators: Theory." Deliverable D3.6 of the EU FP6 project SafetyNet. European Commission, Directorate-General Transport and Energy.

Heinrich, H. W. 1931. *Industrial Accident Prevention: A Scientific Approach*. New York: McGraw-Hill.

Herve, Véronique, and Philippe Lesire. 2017. "Le Véhicule À Conduite Déléguée: Les Enjeux En Matière de Sécurité Routière."

Hollnagel, Erik. 1998. *Cognitive Reliability and Error Analysis Method: CREAM*. 1st ed. Oxford ; New York: Elsevier.

———. 2012. *FRAM, the Functional Resonance Analysis Method: Modelling Complex Socio-Technical Systems*. Farnham, Surrey, UK England ; Burlington, VT: Ashgate.

Hottentot, Chris, Veronique Meines, and Mike Pinkaers. 2015. "Experiments on Autonomous an Automated Driving: An Overview 2015." The Hague: ANWB.

Hughes, B. P., A. Anund, and T. Falkmer. 2015. "System Theory and Safety Models in Swedish, UK, Dutch and Australian Road Safety Strategies." *Accident Analysis & Prevention* 74 (January): 271–78. doi:10.1016/j.aap.2014.07.017.

Hughes, B. P., S. Newstead, A. Anund, C. C. Shu, and T. Falkmer. 2015. "A Review of Models Relevant to Road Safety." *Accident Analysis & Prevention* 74 (January): 250–70. doi:10.1016/j.aap.2014.06.003.

Innamaa, Satu, Scott Smith, and Nobuyuki Uchida. 2016. "A Framework for the Impact Assessment: International Cooperation."

ITF. 2016. *Zero Road Deaths and Serious Injuries*. OECD Publishing. doi:10.1787/9789282108055-en.

Jamson, A. Hamish, Natasha Merat, Oliver M.J. Carsten, and Frank C.H. Lai. 2013. "Behavioural Changes in Drivers Experiencing Highly-Automated Vehicle Control in Varying Traffic

Conditions." *Transportation Research Part C: Emerging Technologies* 30 (May): 116–25. doi:10.1016/j.trc.2013.02.008.

Kaber, David B., and Mica R. Endsley. 1997. "Out-of-the-Loop Performance Problems and the Use of Intermediate Levels of Automation for Improved Control System Functioning and Safety." *Process Safety Progress* 16 (3): 126–131.

Karabatsou, V., M. Pappas, P Van Elslande, K. Fouquet, M. Stanzel, B. Fildes, and R. de Lange. 2007. "A Priori Evaluation of Safety Functions Effectiveness - Methodologies." TRACE Deliverable D4.1.3.

Koopman, Philip, and Michael Wagner. 2016. "Challenges in Autonomous Vehicle Testing and Validation." *SAE International Journal of Transportation Safety* 4 (1): 15–24. doi:10.4271/2016-01-0128.

Kulmala, Risto, Pekka Leviäkangas, Niina Sihvola, Pirkko Rämä, Jonathan Francsics, Ewan Hardman, Simon Ball, et al. 2007. "Final Study Report." CODIA Deliverable.

Kyriakidis, M., J. C. F. de Winter, N. Stanton, T. Bellet, B. van Arem, K. Brookhuis, M. H. Martens, et al. 2017. "A Human Factors Perspective on Automated Driving." *Theoretical Issues in Ergonomics Science*, March, 1–27. doi:10.1080/1463922X.2017.1293187.

Lambert, Fred. 2016. "Understanding the Fatal Tesla Accident on Autopilot and the NHTSA Probe." *Electrek*. July 1. https://electrek.co/2016/07/01/understanding-fatal-tesla-accident-autopilot-nhtsa-probe/.

Larsson, Peter, Sidney W. A. Dekker, and Claes Tingvall. 2010. "The Need for a Systems Theory Approach to Road Safety." *Safety Science*, Scientific Research on Road Safety Management, 48 (9): 1167–74. doi:10.1016/j.ssci.2009.10.006.

Le Coze, Jean-Christophe. 2013. "New Models for New Times. An Anti-Dualist Move." *Safety Science* 59 (November): 200–218. doi:10.1016/j.ssci.2013.05.010.

Leveson, Nancy. 1995. *Safeware: System Safety and Computers*. Addison-Wesley.

———. 2004. "A New Accident Model for Engineering Safer Systems." *Safety Science* 42 (4): 237–70. doi:10.1016/S0925-7535(03)00047-X.

———. 2011. *Engineering a Safer World: Systems Thinking Applied to Safety*. Engineering Systems. Cambridge, Mass: MIT Press.

———. 2015. "A Systems Approach to Risk Management through Leading Safety Indicators." *Reliability Engineering & System Safety* 136: 17–34. doi:10.1016/j.ress.2014.10.008.

———. 2017a. "CAST Analysis of the Shell Moerdijk Accident." Massachusetts Institute of Technology.

———. 2017b. "Rasmussen's Legacy: A Paradigm Change in Engineering for Safety." *Applied Ergonomics* 59 (March): 581–91. doi:10.1016/j.apergo.2016.01.015.

Leveson, Nancy, and John Thomas. 2013. "An STPA Primer."

Ljung Aust, Mikael, Azra Habibovic, Emma Tivesten, Ulrich Sander, Jonas Bärgman, and Engström. 2012. “Manual for DREAM Version 3.2.” Chalmers University of Technology.

Manzie, Chris, Harry Watson, and Saman Halgamuge. 2007. “Fuel Economy Improvements for Urban Driving: Hybrid vs. Intelligent Vehicles.” *Transportation Research Part C: Emerging Technologies* 15 (1): 1–16. doi:10.1016/j.trc.2006.11.003.

Maurer, Markus, J. Christian Gerdes, Barbara Lenz, and Hermann Winner. 2016. *Autonomous Driving*. New York, NY: Springer Berlin Heidelberg.

Merat, N., and A.H. Jamson. 2009. “Is Drivers’ Situation Awareness Influenced by a Fully Automated Driving Scenario?” In *Human Factors: A System View of Human, Technology and Organisation: HFES Europe Chapter*. Soesterberg, the Netherlands.

Merat, Natasha, A. Hamish Jamson, Frank C.H. Lai, Michael Daly, and Oliver M.J. Carsten. 2014. “Transition to Manual: Driver Behaviour When Resuming Control from a Highly Automated Vehicle.” *Transportation Research Part F: Traffic Psychology and Behaviour* 27 (November): 274–82. doi:10.1016/j.trf.2014.09.005.

Merat, Natasha, and Dick de Waard. 2014. “Human Factors Implications of Vehicle Automation: Current Understanding and Future Directions.” *Transportation Research Part F: Traffic Psychology and Behaviour* 27 (November): 193–95. doi:10.1016/j.trf.2014.11.002.

Michon, John A. 1985. “A Critical View of Driver Behavior Models: What Do We Know, What Should We Do?” In *Human Behavior and Traffic Safety*, edited by Leonard Evans and Richard C. Schwing, 485–524. Boston, MA: Springer US. http://link.springer.com/10.1007/978-1-4613-2173-6_19.

Minelli, Simon, Pedram Izadpanah, and Saiedeh Razavi. 2015. “Evaluation of Connected Vehicle Impact on Mobility and Mode Choice.” *Journal of Traffic and Transportation Engineering (English Edition)* 2 (5): 301–12. doi:10.1016/j.jtte.2015.08.002.

Ministère de l’environnement, de l’énergie et de la mer, chargée des relations internationales sur le climat. 2015. *Ordonnance N° 2016-1057 Du 3 Août 2016 Relative À L’expérimentation de Véhicules À Délégation de Conduite Sur Les Voies Publiques*.

Montes, Daniel. 2016. “Using STPA to Inform Developmental Product Testing.” Massachusetts Institute of Technology.

National Transportation Board. 2016. “Preliminary Report.” HWY16FH018.

NHTSA. 2008. “National Motor Vehicle Crash Causation Survey: Report to Congress.” DOT HS 811 059. U.S Department of Transportation, National Highway Traffic Safety Administration.

———. 2013. “Preliminary Statement of Policy Concerning Automated Vehicles.” U.S Department of Transportation, National Highway Traffic Safety Administration.

———. 2014. “Human Factors Evaluation of Level 2 and Level 3 Automated Driving Concepts: Concepts of Operation.” Andrew Marinik, Richard Bishop, Vikki Fitchett, Justin F. Morgan, Tammy E. Trimble, & Myra Blanco.

———. 2015. “Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey.” Traffic Safety Facts DOT HS 812 115. U.S Department of Transportation, National Highway Traffic Safety Administration.

Nilsson, G. 2004. “Traffic Safety Dimensions and the Power Model to Describe the Effect of Speed on Safety.” Bulletin 221. Lund, Sweden: Lund Institute of Technology, Department of Technology and Society.

OECD. 1997. “Road Transport Research: Outlook 2000.” Organisation for Economic Co-operation and Development.

ONISR. 2017. “Les Accidents Corporels de La Circulation 2015.” Recueil de données brutes “Document de travail.” Observatoire National Interministériel de la Sécurité Routière.

Page, Y., C. Rivière, S. Cuny, and T. Zangmeister. 2007. “A Posteriori Evaluation of Safety Functions Effectiveness - Methodologies.” TRACE Deliverable D4.2.1.

Parasuraman, Raja, and Victor Riley. 1997. “Humans and Automation: Use, Misuse, Disuse, Abuse.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39 (2): 230–53. doi:10.1518/001872097778543886.

Pendleton, Scott, Hans Andersen, Xinxin Du, Xiaotong Shen, Malika Meghjani, You Eng, Daniela Rus, and Marcelo Ang. 2017. “Perception, Planning, Control, and Coordination for Autonomous Vehicles.” *Machines* 5 (1): 6. doi:10.3390/machines5010006.

Pillath, Susanne. 2016. “Automated Vehicle in the EU.” European Parliamentary Research Service.

Plummer, Libby. 2016. “Tesla Could Stop You Using Autopilot in Its Cars - but Only If You Take Your Hands off the Wheel.” *Mirror*. http://www.mirror.co.uk/tech/tesla-could-stop-you-using-8734846.

Rasmussen & Svedung. 2000. *Proactive Risk Management in a Dynamic Society.* [S.l.]: Swedish Rescue Services A.

Rasmussen, J. 1997. “Risk Management in a Dynamic Society: A Modelling Problem.” *Safety Science* 27 (2): 183–213. doi:10.1016/S0925-7535(97)00052-0.

Rasmussen, Jens. 1986. “A Framework for Cognitive Task Analysis in Systems Design.” In *Intelligent Decision Support in Process Environments*, edited by Erik Hollnagel, Giuseppe Mancini, and David D. Woods, 21:175–96. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-642-50329-0_12.

Rau, Paul, Mikio Yanagisawa, and Wassim G Najm. 2015. “Target Crash Population of Automated Vehicles.” In .

Reason, J. T. 1990. *Human Error*. Cambridge [England] ; New York: Cambridge University Press.

———. 1997. *Managing the Risks of Organizational Accidents*. Aldershot, Hants, England ; Brookfield, Vt., USA: Ashgate.

———. 2008. *The Human Contribution: Unsafe Acts, Accidents and Heroic Recoveries*. Farnham, England ; Burlington, VT: Ashgate.

Risto Kulmala. 2010. “Ex-Ante Assessment of the Safety Effects of Intelligent Transport Systems.” *Accident; Analysis and Prevention* 42 (4): 1359–69. doi:10.1016/j.aap.2010.03.001.

SAE International. 2012. “Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems.”

———. 2015. “Guidelines for Safe on-Road Testing of SAE Level 3, 4, and 5 Prototype AutomatedDriving Systems (ADS).” SurfaceVehicle Informatin Report J3018.

———. 2016. “Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems.” Surfce Vehicle Information Report.

Sagberg, Fridulv, and Transportøkonomisk institutt (Norway). 2008. *A methodological study of the Driving Reliability and Error Analysis Method (DREAM)*. Oslo: Transportøkonomisk institutt.

Salmon, Paul M., and Michael G. Lenné. 2015. “Miles Away or Just around the Corner? Systems Thinking in Road Safety Research and Practice.” *Accident Analysis & Prevention* 74 (January): 243–49. doi:10.1016/j.aap.2014.08.001.

Salmon, Paul M., Rod McClure, and Neville A. Stanton. 2012. “Road Transport in Drift? Applying Contemporary Systems Thinking to Road Safety.” *Safety Science* 50 (9): 1829–38. doi:10.1016/j.ssci.2012.04.011.

Salmon, Paul M., Gemma J.M. Read, and Nicholas J. Stevens. 2016. “Who Is in Control of Road Safety? A STAMP Control Structure Analysis of the Road Transport System in Queensland, Australia.” *Accident Analysis & Prevention* 96 (November): 140–51. doi:10.1016/j.aap.2016.05.025.

Scott-Parker, B., N. Goode, and P. Salmon. 2015. “The Driver, the Road, the Rules … and the Rest? A Systems-Based Approach to Young Driver Road Safety.” *Accident Analysis & Prevention* 74 (January): 297–305. doi:10.1016/j.aap.2014.01.027.

Silla, Anne, Pirkko Rämä, Lars Leden, Martijn van Noort, Janiek de Kruijff, Daniel Bell, Andrew Morris, Graham Hancox, and Johan Scholliers. 2017. “Quantifying the Effectiveness of ITS in Improving Safety of VRUs.” *IET Intelligent Transport Systems* 11 (3): 164–72. doi:10.1049/iet-its.2016.0024.

Singhvi, Anjali, and Karl Russell. 2016. “Inside the Self-Driving Tesla Fatal Accident.” *The New York Times*. July 26. https://www.nytimes.com/interactive/2016/07/01/business/inside-tesla-accident.html?_r=0.

Smith, Bryant Walker. 2016. “Automated Driving and Product Liability.” SSRN Scholarly Paper ID 2923240. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2923240.

Smith, Scott, Jeffrey Bellone, Stephen Bransfield, Amy Ingles, George Noel, Erin Reed, and Mikio Yanagisawa. 2015. “Benefits Estimation Framework for Automated Vehicle Operations.” DOT-VNTSC-FHWA-15-12.

Suo, Dajiang, Sarra Yako, Mathew Boesch, and Kyle Post. 2017. “Integrating STPA into ISO 26262 Process for Requirement Development.” In . doi:10.4271/2017-01-0058.

The Tesla Team. 2016. “Upgrading Autopilot: Seeing the World in Radar.” *Tesla*. September 11. https://www.tesla.com/blog/upgrading-autopilot-seeing-world-radar.

Tingvall, C. 1995. “The Zero Vision; A Road Transport System Free from Serious Health Losses.” In *Transportation, Traffic Safety and Health*, 37–57. Göteborg, Sweden: Holst, H. von, Nygren, Å., Thord,.

Treat, J.R. 1977. “Tri-Level Study of the Causes of Traffic Accidents: An Overview of Final Results.” *Proceedings: American Association for Automotive Medicine Annual Conference* 21: 391–403.

UK Department of Transport. 2015a. “The Pathway to Driverless Cars: A Code of Practice for Testing.”

———. 2015b. “The Pathway to Driverless Cars: Summary Report and Action Plan.”

UN. 1968. “Convention on the Law of Treaties: Convention on Road Traffic.” In , 2042:17. Vienna: United Nations.

Underwood, Peter. 2013. “Examining the Systemic Accident Analysis Research-Practice Gap.” Loughborough University.

Underwood, Peter, and Patrick Waterson. 2013. “Systemic Accident Analysis: Examining the Gap between Research and Practice.” *Accident Analysis & Prevention* 55 (June): 154–64. doi:10.1016/j.aap.2013.02.041.

UNECE. 2016. “UNECE Paves the Way for Automated Driving by Updating UN International Convention.” http://www.unece.org/info/media/presscurrent-press-h/transport/2016/unece-paves-the-way-for-automated-driving-by-updating-un-international-convention/doc.html.

Van Eikema Hommes, Qi. 2012. “Applying STPA to Automotive Adaptive Cruise Control System.”

Van Elslande, Pierre. 2000a. “L’erreur humaine dans les scénarios d’accident cause ou conséquence?Human error in accident scenarios: cause or consequence?” *Recherche - Transports - Sécurité* 66 (January): 7–31. doi:10.1016/S0761-8980(00)90002-5.

———. 2000b. “L’erreur humaine dans les scénarios d’accident cause ou conséquence?Human error in accident scenarios: cause or consequence?” *Recherche - Transports - Sécurité* 66 (January): 7–31. doi:10.1016/S0761-8980(00)90002-5.

Van Elslande, Pierre, and Lydie Alberton. 1997. *Scénarios-types de production de l’erreur humaine dans l’accident de la route: problématique et analyse qualitative*. Arcueil: INRETS.

Van Elslande, Pierre, and Katel Fouquet. 2007. “D5.1 Analyzing ‘Human Functional Failures’ in Road Accidents.” TRACE projet.

Vlahogianni, Eleni I., George Yannis, and John C. Golias. 2012. “Overview of Critical Risk Factors in Power-Two-Wheeler Safety.” *Accident Analysis & Prevention* 49 (November): 12–22. doi:10.1016/j.aap.2012.04.009.

Waterson, Patrick, Jean-Christophe Le Coze, and Henning Boje Andersen. 2017. “Recurring Themes in the Legacy of Jens Rasmussen.” *Applied Ergonomics* 59 (March): 471–82. doi:10.1016/j.apergo.2016.10.002.

Wegman, Fred, Letty Aarts, and Charlotte Bax. 2008. “Advancing Sustainable Safety.” *Safety Science* 46 (2): 323–43. doi:10.1016/j.ssci.2007.06.013.

Winter, Joost C.F. de, Riender Happee, Marieke H. Martens, and Neville A. Stanton. 2014. “Effects of Adaptive Cruise Control and Highly Automated Driving on Workload and Situation Awareness: A Review of the Empirical Evidence.” *Transportation Research Part F: Traffic Psychology and Behaviour* 27 (November): 196–217. doi:10.1016/j.trf.2014.06.016.

World Health Organization. 2015. *Global Status Report on Road Safety 2015: Supporting a Decade of Action*. Geneva, Switzerland: WHO.

Young, Kristie L., and Paul M. Salmon. 2015. “Sharing the Responsibility for Driver Distraction across Road Transport Systems: A Systems Approach to the Management of Distracted Driving.” *Accident Analysis & Prevention* 74 (January): 350–59. doi:10.1016/j.aap.2014.03.017.

Young, M. S., and N. A. Stanton. 2007. “What’s Skill Got to Do with It? Vehicle Automation and Driver Mental Workload.” *Ergonomics* 50 (8): 1324–39. doi:10.1080/00140130701318855.

Appendix A: STPA results (chapter 3)

Table 41 – STPA results for category 1

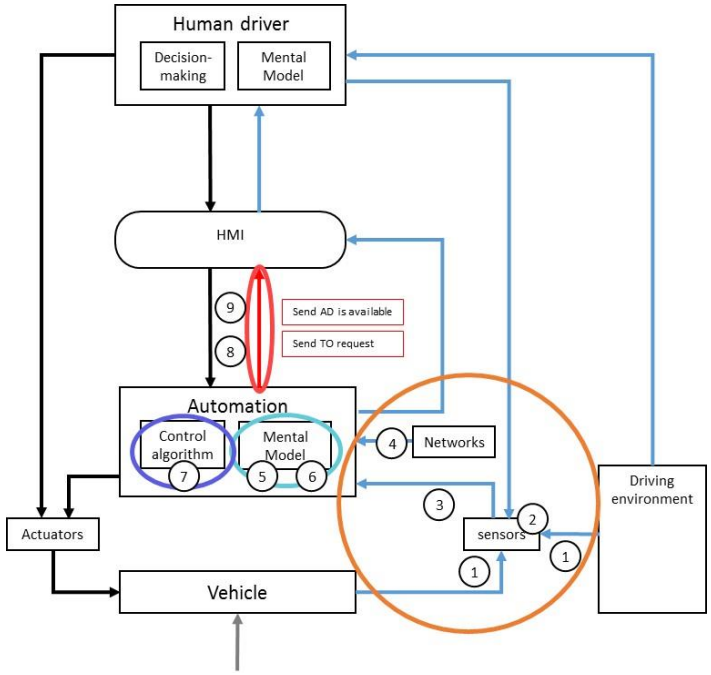
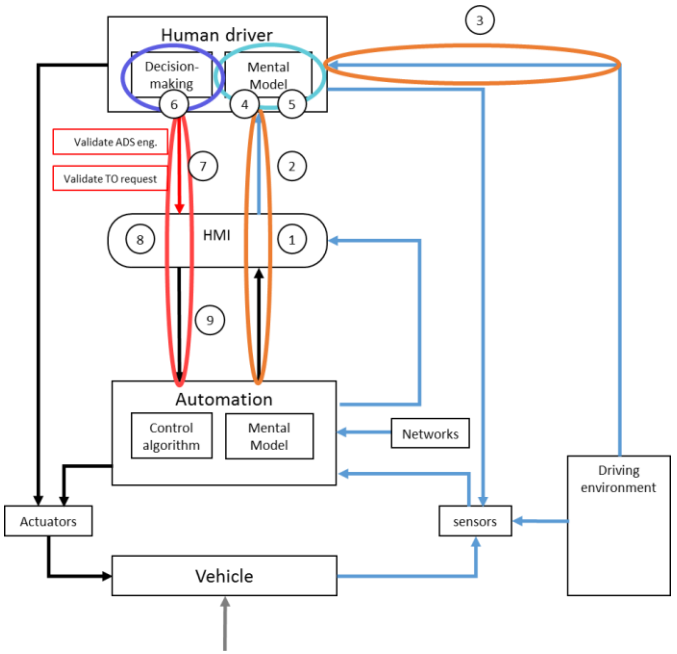
CATEGORY 1: Automation sends feedback to influence a transition			
STPA step 1		**UCAs translated into safety requirements**	
Unsafe control actions		**Safety requirements**	
UCA-3: Automation sends ADS availability notification when ADS is not available		**SR-3**: Automation must not send ADS availability notification when ADS is not available	
UCA-16: Automation does not send takeover request when the ADS conditions are no longer met (end mode type 2)		**SR-16**: Automation must send takeover request when the ADS conditions are no longer met (end mode type 2)	
UCA-17: Automation does not send takeover request when the ADS compatible road comes to an end e.g. highway exit (end mode type 1)		**SR-17:** Automation must send takeover request when the ADS compatible road comes to an end e.g. highway exit (end mode type 1)	
STPA step 2			
Control structure	**Class**	**Control flaw**	**Refined safety requirement**
	Feedback and inputs	Measurement inaccuracies, feedback delays or no information measured by vehicle sensors (1)	**RSR-1:** Vehicle sensors must take accurate on time measures on the necessary feedback to determine that ADS is available and that a takeover request is needed
			RSR-2: Automation must detect when vehicle sensors are providing inaccurate measures with delays of TBD, on the necessary feedback to determine that ADS is available and that a takeover request is needed
		Inadequate sensor operation (2)	**RSR-3:** Vehicle sensors that measure the necessary feedback to determine that ADS is available and that a takeover request is needed, must have an adequate operation
			RSR-4: Automation must detect when the vehicle sensors that provide the necessary feedback to determine that ADS is available and that a takeover request is needed, have an inadequate operation
		Missing or inadequate feedback on ADS conditions sent by vehicle sensors (3)	**RSR-5:** Vehicle sensors must provide adequate feedback on the necessary information to determine that ADS is available and that a takeover request is needed, have an inadequate operation
		Missing or inadequate external information on ADS conditions (4)	**RSR-6:** External information (e.g. networks) must provide adequate information on the necessary feedback to determine that ADS is available and that a takeover request is needed, have an inadequate operation
	Model	Inadequate model of the state of ADS conditions (5)	**RSR-7:** Automation must have an adequate model of ADS availability conditions and an adequate model of ADS conditions to continue on automated driving
		Inadequate model of ADS conditions (6)	**RSR-8:** The software requirements and feedback inputs included in automation's model must enable automation to adequately assess the state of ADS conditions
	Decision-making	Inadequate control algorithm (7)	**RSR-9:** Automation's control algorithm must not generate ADS availability notification when the model indicates ADS is not available, and must generate takeover requests when the ADS conditions are no longer met
	Action execution	Missing control action (8)	**RSR-10:** Automation must ensure that the actions generated by the control algorithm to send the ADS availability notification and the takeover requests to the HMI, are executed
		Delayed operation (9)	**RSR-11:** Automation must ensure that the actions generated by the control algorithm to send the ADS availability notification and the takeover requests to the HMI are sent with a maximal delay of TBD

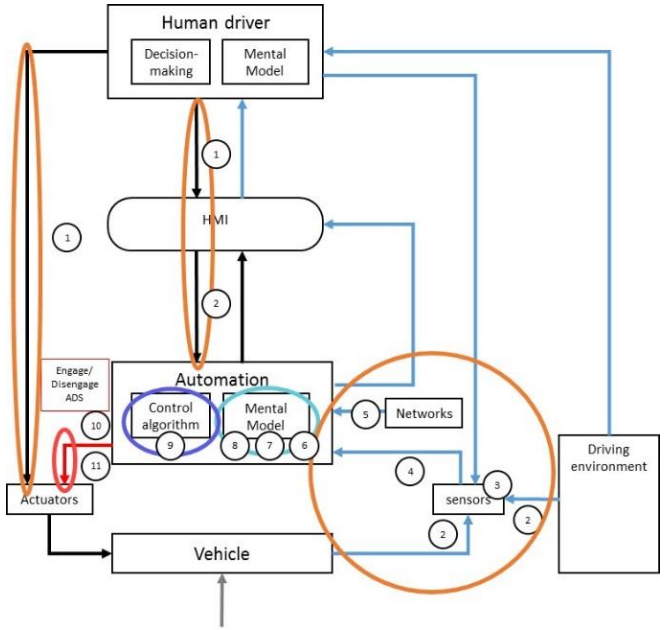
Table 42 – STPA results for category 2

<table>
<tr><th colspan="4">CAREGORY 2: Driver responds to feedback sent by automation</th></tr>
<tr><th colspan="2">STPA step 1</th><th colspan="2">UCAs translated into safety requirements</th></tr>
<tr><td colspan="2">Unsafe control actions</td><td colspan="2">Safety requirements</td></tr>
<tr><td colspan="2">UCA-4: Driver provides ADS validation when it is inappropriate to engage ADS</td><td colspan="2">SR-4: Driver must not provide ADS validation when it is inappropriate to engage ADS</td></tr>
<tr><td colspan="2">UCA-18: Driver does not validate the takeover request when automation sends the takeover request</td><td colspan="2">SR-18: Driver must validate the takeover request when automation sends the takeover request</td></tr>
<tr><td colspan="2">UCA-19: Driver validates takeover request and puts the vehicle in an unsafe situation</td><td colspan="2">SR-19: Driver must not put the vehicle in an unsafe situation after the validation of the takeover request</td></tr>
<tr><th colspan="4">STPA step 2</th></tr>
<tr><th>Control structure</th><th>Class</th><th>Control flaw</th><th>Refined safety requirement</th></tr>
<tr><td rowspan="11"> 
</td><td rowspan="8">Feedback and inputs</td><td rowspan="3">Inadequate or missing feedback provided by the HMI (1)</td><td>RSR-12: There must be an adequate communication between automation and the HMI, and an adequate HMI operation that enables to display the feedback provided by automation on ADS availability notification and takeover requests.</td></tr>
<tr><td rowspan="3">Inadequate human perception on the HMI (2)</td><td>RSR-13: The HMI must provide adequate feedback to the driver on ADS availability notification and takeover requests.</td></tr>
<tr><td>RSR-14: The mental model of the driver must include the procedures and knowledge necessary to understand the feedback provided by the HMI.</td></tr>
<tr><td>RSR-15: The driver must value being receptive to the feedback provided by the HMI</td></tr>
<tr><td rowspan="3">Inadequate human perception on the traffic environment (3)</td><td>RSR-16: The driver must be able to perceive and detect the aspects that make it inappropriate to engage the ADS</td></tr>
<tr><td>RSR-17: The takeover procedures must enable the driver to perceive the traffic environment before the validation of the takeover request</td></tr>
<tr><td rowspan="3">Model</td><td rowspan="3">Inadequate model of takeover request (4)</td><td>RSR-18: The mental model of the driver must include knowledge on the takeover procedures</td></tr>
<tr><td>RSR-19: The procedures to validate a takeover request must be intuitive and easy to perform by the driver</td></tr>
<tr><td>RSR-20: The HMI must provide adequate feedback to the driver on the steps to validate a takeover request</td></tr>
<tr><td rowspan="2">Inadequate model of the driving environment (5)</td><td>RSR-21: The mental model of the driver must include the situations when it is inappropriate to engage ADS</td></tr>
<tr><td>RSR-22: The driver must have an adequate model of the traffic environment before the validation of the ADS engagement and takeover requests</td></tr>
</table>

STPA step 2			
Control structure	**Class**	**Control flaw**	**Refined safety requirement**
	Decision-making	Inadequate control algorithm (6)	**RSR-23:** The mental model of the driver must include safety values that encourage an adequate decision-making process regarding ADS engagement and takeover request validation
	Action execution	Inappropriate control action (7)	**RSR-24:** The procedures and commands to validate ADS engagement and takeover requests must limit unintended validations
		Missing control action (7) and (8)	**RSR-25:** The mental model of the driver must include the location of the validation commands, the sequences, order, etc.
			RSR-26: The design of the validation commands and the HMI display information with takeover request must assist the driver to safely validate takeover requests
		Inadequate actuator operation and communication (8) and (9)	**RSR-27:** The HMI commands must have an adequate operation and there must be an adequate communication between the HMI and automation, which ensures the actions provided by the driver reach automation

Table 43 – STPA results for category 3

CATEGORY 3: Automation engages/disengages ADS	
STPA step 1	**UCAs translated into safety requirements**
Unsafe control actions	**Safety requirements**
UCA-5: Automation does not engage ADS when the driver validates ADS engagement	**SR-5:** Automation must engage ADS when the driver validates ADS engagement
UCA-6: Automation provides ADS engagement when ADS conditions are not met	**SR-6:** Automation must not provide ADS engagement when ADS conditions are not met
UCA-7: Automation provides ADS engagement when the driver does not validate ADS engagement	**SR-7:** Automation must not provide ADS engagement when the driver does not validate ADS engagement
UCA-20: Automation does not disengage ADS when the driver validates a takeover request	**SR-20:** Automation must disengage ADS when the driver validates a takeover request
UCA-21: Automation disengages ADS when the driver has not validated a takeover request	**SR-21:** Automation must not disengage ADS when the driver has not validated a takeover request
UCA-29: Automation does not disengage ADS when the driver provides ADS disengagement (via ADS disengagement commands or control override)	**SR-29:** Automation must disengage ADS when the driver provides ADS disengagement (via ADS disengagement commands or control override)
UCA-30: Automation disengages ADS when the driver has not provided ADS disengagement (via ADS disengagement commands or control override).	**SR-30:** Automation must not disengage ADS when the driver has not provided ADS disengagement (via ADS disengagement commands or control override).

STPA step 2			
Control structure	**Class**	**Control flaw**	**Refined safety requirement**
	Feedback and inputs	Missing or inadequate feedback on driver's actions (1)	**RSR-28:** The HMI commands, pedals and steering wheel must be reliable and provide on-time feedback on driver's ADS engagement/disengagement and takeover validation
			RSR-29: Automation must receive adequate feedback on driver's actions (engagement and disengagements)
		Measurement inaccuracies, feedback delays or no information measured by vehicle sensors (2)	**RSR-30:** Vehicle sensors must take accurate measures on ADS conditions with a maximal delay of TBD
			RSR-31: Automation must detect when the sensors that measure ADS conditions measurements have an inadequate operation
		Inadequate sensor operation (3)	**RSR-32:** Vehicle sensors that measure the necessary feedback to evaluate ADS conditions, must have an adequate operation
			RSR-33: Automation must detect when the vehicle sensors that provide the necessary feedback to evaluate ADS conditions, have an inadequate operation.
		Missing or inadequate feedback on sent by vehicle sensors (4)	**RSR-34:** The feedback provided by vehicle sensors on ADS conditions must be adequate
		Missing or inadequate external information on ADS conditions (5)	**RSR-35:** The feedback provided by external information on ADS conditions must be adequate
	Model	Inadequate model of the status of ADS engagement/ disengagement (6)	**RSR-36:** Automation must have an adequate model of the status of ADS engagement/disengagement
		Inadequate model of the state of ADS conditions (7)	**RSR-37:** Automation must have an adequate model of ADS conditions (when they are not met)
		Inadequate model of the status of a takeover request (8)	**RSR-38:** Automation must have an adequate model of the status of a takeover request (validated or not validated)

STPA step 2			
Control structure	**Class**	**Control flaw**	**Refined safety requirement**
	Decision-making	Inadequate control algorithm (engagement) (9)	**RSR-39:** Automation's control algorithm must not generate ADS engagement when the driver has not validated engagement and when ADS conditions are not met, and must generate engagement when driver validates engagement
	Action execution	Inappropriate control actions (10)	**RSR-40:** Automation must ensure that the actions generated by the control algorithm related to ADS engagement/ disengagement are appropriate
		Delayed operation (11)	**RSR-41:** Automation must ensure that the actions generated by the control algorithm related to ADS engagement/ disengagement are provided with a maximal delay of TBD

Table 44 – STPA results for category 4

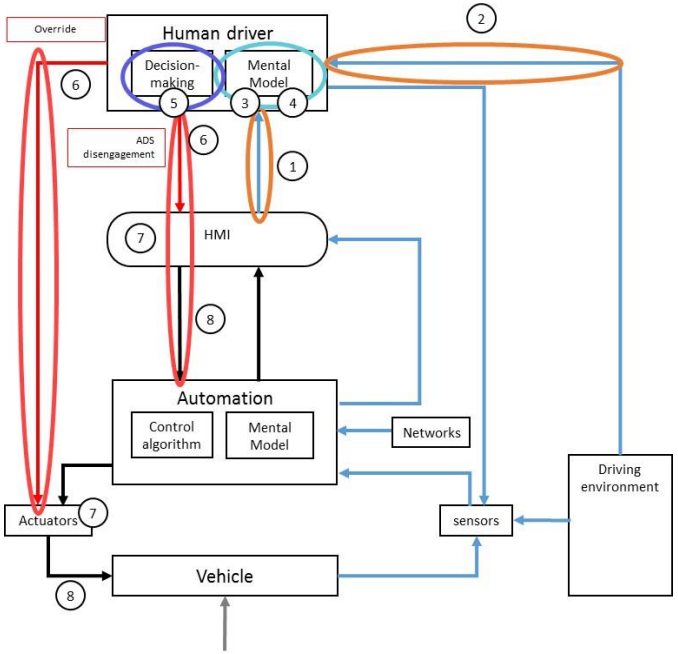
CAREGORY 4: Driver disengages ADS (on driver's request)			
STPA step 1			
Unsafe control actions		**Safety requirements**	
UCA-12: Driver disengages the ADS (on driver's request) and puts the vehicle in an unsafe situation		**SR-12**: The driver must not put the vehicle in an unsafe situation when s/he disengages the ADS	
STPA step 2			
Control structure	**Class**	**Control flaw**	**Refined safety requirement**
	Feedback and inputs	Inadequate human perception on the HMI (1)	**RSR-42:** The driver must perceive the HMI information regarding ADS disengagement
		Inadequate human perception on the traffic environment (2)	**RSR-43:** The driver must perceive the driving environment before disengaging the ADS
	Model	Inadequate model of ADS disengagement procedure (3)	**RSR-44:** The mental model of the driver must include knowledge on ADS disengagement procedure and the HMI (sequences, buttons, HMI displays, etc.)
		Inadequate model of the driving environment (4)	**RSR-45:** The driver must have an adequate model of the traffic environment before the validation of the ADS disengagement
	Decision-making	Inadequate control algorithm (5)	**RSR-46:** The mental model of the driver must include safety values that encourage an adequate decision-making process regarding ADS disengagement
	Action execution	Inappropriate control action (6)	**RSR-47:** The driver must not provide unintended ADS disengagement
			RSR-48: The design of the ADS system must limit unintended ADS disengagements
		Inadequate actuator operation and communication (7) and (8)	**RSR-49:** The HMI commands and vehicle actuators that enable ADS disengagement must have an adequate operation and communication

Table 45 – STPA results for category 5

CATEGORY 5: Automation provides control of the vehicle			
STPA step 1			
Unsafe control actions		**Safety requirements**	
UCA-2: Automation provides control of the vehicle during manual driving		**SR-2**: Automation must not provide control of the vehicle during manual driving	
UCA-8: Automation does not provide control of the vehicle after ADS engagement		**SR-8:** Automation must provide control of the vehicle after ADS engagement	
UCA-9: Automation provides inadequate control of the vehicle after ADS engagement		**SR-9:** Automation must provide adequate control of the vehicle after ADS engagement	
UCA-13: Automation provides control of the vehicle after ADS conditions are no longer met		**SR-13:** Automation must not provide control of the vehicle after ADS conditions are no longer met	
UCA-14: Automation follows traffic rules and/or social norms in an inadequate fashion during automated driving.		**SR-14:** Automation must follow traffic rules and/or social norms in an adequate fashion during automated driving.	
UCA-15: Automation follows traffic rules and/or social norms during automated driving and puts the vehicle in an unsafe situation		**SR-15:** Automation must not put the vehicle in an unsafe situation when automation follows traffic rules and/or social norms during automated	
UCA-22: Automation does not release control of the vehicle when the driver validates a takeover request		**SR-22:** Automation must release control of the vehicle when the driver validates a takeover request	
UCA-31: Automation does not release control of the vehicle when the driver overrides ADS or provides ADS disengagement.		**SR-31:** Automation must release control of the vehicle when the driver overrides ADS or provides ADS disengagement.	
UCA-25: Automation does not provide minimal risk maneuver when the driver does not respond to the takeover request (end mode type 1 and end mode type 2)		**SR-25:** Automation must minimal risk maneuver when the driver does not respond to the takeover request (end mode type 1 and end mode type 2)	
UCA-26: Automation does not provide minimal risk maneuver when automation can no longer assure safe operation (end type mode 3)		**SR-26:** Automation must provide minimal risk maneuver when automation can no longer assure safe operation (end type mode 3)	
UCA-27: Automation provides minimal risk maneuver and puts the vehicle in an unsafe situation		**SR-27:** Automation must not put the vehicle in an unsafe situation when automation provides a minimal risk maneuver	
STPA step 2			
Control structure	**Class**	**Control flaw**	**Refined safety requirement**
	Feedback and inputs	Inadequate or missing feedback on driver's actions (1)	**RSR-50:** The HMI commands and vehicle actuators must provide adequate, on-time feedback on driver's actions
		Incorrect feedback on driving mode status (2)	**RSR-51:** Automation must receive adequate feedback on the driving mode status (manual or automated driving mode)
		Measurement inaccuracies, feedback delays or no information measured by vehicle sensors (3)	**RSR-52:** Sensors must take accurate on-time measures on ADS conditions and the traffic environment
			RSR-53: Automation must detect when sensors are providing inaccurate measures on ADS conditions and the traffic environment or measures with feedback delays
		Inadequate sensor operation (4)	**RSR-54:** The vehicle sensors that take measures on ADS conditions and the traffic environment must have an adequate operation
			RSR-55: Automation must detect when the sensors that take measures on ADS conditions and the driving environment have an inadequate operation

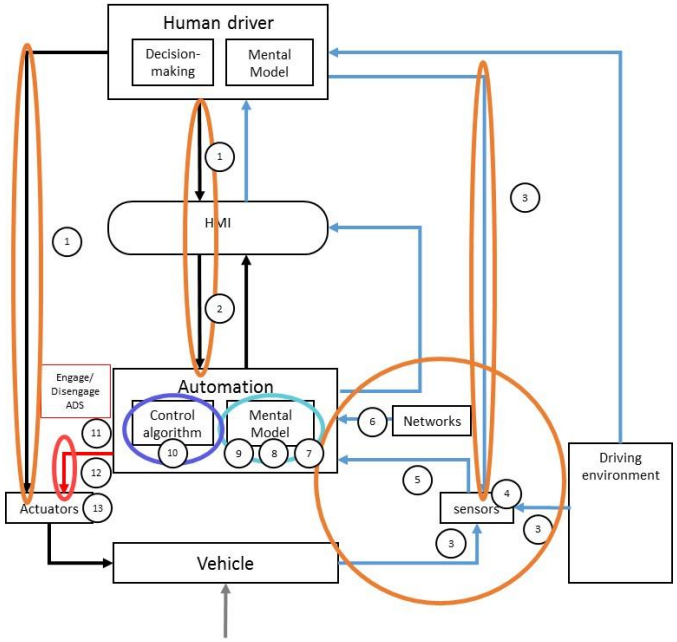
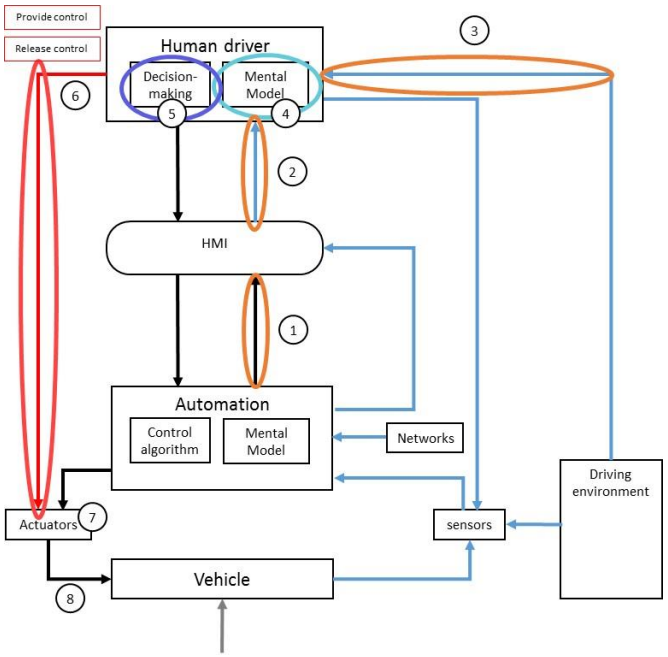
STPA step 2			
Control structure	**Class**	**Control flaw**	**Refined safety requirement**
	Feedback and inputs	Missing or inadequate feedback provided by vehicle sensors (5)	**RSR-56:** The feedback provided by vehicle sensors on ADS conditions driver monitoring, and the traffic environment, must be adequate
		Missing or inadequate external information on ADS conditions (6)	**RSR-57:** The feedback provided by external information on ADS conditions must be adequate
	Model	Inadequate model of the state of ADS (7)	**RSR-58:** The model of the state of ADS (engaged or disengaged) must be adequate
		Inadequate model of driver (8)	**RSR-59:** The model of human driver must be adequate
		Inadequate model of the state of the traffic environment, traffic rules and social norms (9)	**RSR-60:** The design assumptions must enable automation to have an adequate representation of the traffic environment, traffic rules and social norms
			RSR-61: Automation must have a prioritization for safety, traffic rules and social norms.
	Decision-making	Inadequate control algorithm (10)	**RSR-62:** Automation's control algorithm must not generate actions during manual driving mode
			RSR-63: Automation's control algorithm must not generate appropriate actions to control the vehicle and comply with traffic rules and social norms.
			RSR-64: Automation's control algorithm must release control of the vehicle after ADS disengagement.
			RSR-65: Automation's control algorithm must generate appropriate actions to perform a minimal risk maneuver.
	Action execution	Inappropriate control actions (11)	**RSR-66:** Automation must implement in an appropriate fashion the control actions that allow to control the vehicle, comply with rules and norms, release control, and perform minimal risk maneuver.
		Delayed operation (12)	**RSR-67:** The control actions for vehicle control, rule and norm compliance, release of control and the minimal risk maneuver, must be sent with maximal delay of TBD
		Inadequate actuator operation (13)	**RSR-68:** The actuators that enable the control actions for vehicle control, rule and norm compliance, release of control and the minimal risk maneuver, must have an adequate operation.

Table 46 – STPA results for category 6

<table>
<tr><th colspan="4">CAREGORY 6: Driver provides control of the vehicle</th></tr>
<tr><th colspan="4">STPA step 1</th></tr>
<tr><th colspan="2">Unsafe control actions</th><th colspan="2">Safety requirements</th></tr>
<tr><td colspan="2">UCA-1: Driver provides inadequate control of the vehicle during manual driving</td><td colspan="2">UCA-1: Driver provides inadequate control of the vehicle during manual driving</td></tr>
<tr><td colspan="2">UCA-23: Driver does not provide control of the vehicle after the validation of a takeover request</td><td colspan="2">UCA-23: Driver does not provide control of the vehicle after the validation of a takeover request</td></tr>
<tr><td colspan="2">UCA-24: Driver provides inadequate control of the vehicle during manual driving</td><td colspan="2">UCA-24: Driver provides inadequate control of the vehicle during manual driving</td></tr>
<tr><td colspan="2">UCA-28: Driver provides inadequate control of the vehicle after a minimal risk maneuver</td><td colspan="2">UCA-28: Driver provides inadequate control of the vehicle after a minimal risk maneuver</td></tr>
<tr><td colspan="2">UCA-32: Driver provides inadequate control of the vehicle after ADS disengagement</td><td colspan="2">UCA-32: Driver provides inadequate control of the vehicle after ADS disengagement</td></tr>
<tr><td colspan="2">UCA-10: Driver does not release control of the vehicle after ADS engagement and puts the vehicle in an unsafe situation.</td><td colspan="2">UCA-10: Driver does not release control of the vehicle after ADS engagement and puts the vehicle in an unsafe situation.</td></tr>
<tr><td colspan="2">UCA-11: Driver releases control of the vehicle too soon before the ADS is engaged.</td><td colspan="2">UCA-11: Driver releases control of the vehicle too soon before the ADS is engaged.</td></tr>
<tr><th colspan="4">STPA step 2</th></tr>
<tr><th>Control structure</th><th>Class</th><th>Control flaw</th><th>Refined safety requirement</th></tr>
<tr><td rowspan="7"> 
</td><td rowspan="5">Feedback and inputs</td><td>Inadequate or missing feedback provided by the HMI (1)</td><td>RSR-69: There must be an adequate communication between automation and the HMI, and an adequate HMI operation that enables to display the feedback provided by automation on ADS status, takeover requests and minimal risk maneuver</td></tr>
<tr><td rowspan="3">Inadequate human perception on the HMI (2)</td><td>RSR-70: The HMI must provide adequate feedback to the driver on ADS status, takeover requests and minimal risk maneuver</td></tr>
<tr><td>RSR-71 The mental model of the driver must include the procedures and knowledge necessary to understand the feedback provided by the HMI on ADS status, takeover request and minimal risk maneuvers</td></tr>
<tr><td>RSR-72: The driver must value being receptive to the feedback provided by the HMI</td></tr>
<tr><td>Inadequate human perception on the traffic environment (3)</td><td>RSR-73: The takeover procedures must enable the driver to perceive the traffic environment before the validation of the takeover request</td></tr>
<tr><td rowspan="3">Model</td><td rowspan="3">Inadequate model of the driving environment (4)</td><td>RSR-74: The mental model of the driver must include the procedures to validate ADS engagement and takeover requests, disengage the ADS and release control of the vehicle</td></tr>
<tr><td>RSR-75: The ADS procedures must enable the driver to safely validate ADS engagement and takeover requests, disengage the ADS and release control of the vehicle</td></tr>
<tr><td>RSR-76: The HMI must provide feedback to assist the driver in the validation of ADS engagement and takeover requests, disengagement of the ADS and release control of the vehicle</td></tr>
</table>

STPA step 2			
Control structure	**Class**	**Control flaw**	**Refined safety requirement**
	Decision-making	Inadequate control algorithm (5)	**RSR-77:** The mental model of the driver must include safety values that encourage an adequate decision-making process during automated driving
	Action execution	Inappropriate control action (6)	**RSR-78:** The driver must be aware of the vehicle cockpit configuration, command locations, sequences, etc. necessary for the validation of ADS engagement and takeover requests and disengagement of the ADS
			RSR-79: The ADS procedures must enable the driver to safely release control of the vehicle and resume manual driving
		Inadequate actuator operation and communication (7) and (8)	**RSR-80:** The actuators and commands to implement ADS engagement validation, takeover validation, ADS disengagement and control of the vehicle, must have an adequate operation

Appendix B: Results of the STPA on the vehicle trial process (chapter 4)

Table 47 – STPA on the vehicle trial process

Controller	STPA step 1	Safety Requirement	STPA step 2		
	Unsafe control Action		Control flaw	Scenario	Refined Safety Requirement
Government	**UCA-1:** The government establishes inadequate regulations for vehicle trials	**SR-1:** The government must establish adequate regulations for vehicle trials	Inadequate mental model: Relevance of existing regulation	The government does not establish adequate regulations for vehicle trials because they believe that the existing regulations are enough to regulate trial safety	**RSR-1.1:** The government must have an adequate model of the relevance of the exiting regulations and the need for new regulations
			Inadequate mental model: Vehicle technology	The government does not establish adequate regulations for vehicle trials because they have an inadequate understanding of vehicle technology	**RSR-1.2:** The government must have an adequate model of the vehicle technology being tested
	UCA-2: The government authorizes unsafe vehicle trials	**SR-2:** The government must not authorize unsafe vehicle trials	Inadequate mental model: Safety of the vehicle trial	The government authorizes an unsafe vehicle trial because they are not aware that the trial is unsafe	**RSR-2.1:** The government must have an adequate model of the vehicle trial
					RSR-2.2: The trial manager must provide adequate feedback in the dossier for a trial authorization request
Funding agencies	**UCA-3:** The funding agencies provide funding for unsafe vehicle trials	**SR-3:** The funding agencies must not provide funding for unsafe vehicle trials	Inadequate mental model: Safety of the vehicle trial	The funding agencies provide funding for an unsafe vehicle trials because they are not aware that the trial is unsafe	**RSR-3.1:** The funding agencies must have an adequate model of the vehicle trial
					RSR-3.2: The trial manager must provide adequate feedback in trial proposal
	UCA-4: The funding agencies set trial conditions that contribute to unsafe vehicle trials	**SR-4:** The funding agencies must not set trial conditions that contribute to unsafe vehicle trials	Inadequate mental model: Safety of trial conditions	The funding agencies set trial conditions that contribute to unsafe vehicle trials because they believe the requirements are safe	**RSR-4.1:** The funding agencies must have an adequate model of safety of trial conditions that they set
Company Management	**UCA-5:** The company management defines an inadequate roadmap that leads to the development of unsafe vehicle technologies and unsafe trials	**SR-5:** The company management must define an adequate roadmap that facilitates the development of safe vehicle technologies and safe trials	Inadequate mental model: Need for a clear and understandable roadmap	The company management defines an inadequate roadmap because they consider that the roadmap does not need to be clear an understandable to all employees	**RSR-5.1:** The company management must define a clear and understandable roadmap and diffuse it to all employees
			Inadequate mental model: Safety of the roadmap	The company management defines an inadequate roadmap because they have an incorrect model of the roadmap's safety (they believe that it is safe when it is not)	**RSR-5.2:** The company management's model must include the knowledge and information necessary to assess roadmap's safety
					RSR-5.3: The lower levels of the company must provide company management with adequate feedback on hazards associated to vehicle technologies and vehicle trials
			Inadequate mental model: Safety vs innovation	The company management defines an inadequate roadmap because they believe that safety can be compromised for the sake of innovation	**RSR-5.4:** The company management must evaluate their safety value regarding the place of automation relative to innovation

Controller	STPA step 1	Safety Requirement	STPA step 2		
	Unsafe control Action		Control flaw	Scenario	Refined Safety Requirement
Company Management	**UCA-5:** The company management defines an inadequate roadmap that leads to the development of unsafe vehicle technologies and unsafe trials	**SR-5:** The company management must define an adequate roadmap that facilitates the development of safe vehicle technologies and safe trials	Inadequate mental model: Roadmap dissemination	The company management defines an inadequate roadmap because even if the roadmap is clear and safe, they do not consider it important to disseminate it to all employees concerned by it	**RSR-5.5:** The company management must establish strategies to disseminate the roadmap to all employees concerned by it
	UCA-6: The company management provides inadequate standards and resources for vehicle trials	**SR-6:** The company management must provide adequate standards and resources for vehicle trials	Inadequate mental model: Relevance of current standards and resources	The company management does not provide adequate standards and resources for vehicle trials because they believe that the existing standards and resources are enough to ensure safe trials	**RSR-6.1-:** The company management must have an adequate model of the relevance of the existing standards and resources, and the need for new regulation for new ones
			Inadequate mental model: Vehicle technologies and vehicle trials	The company management does not provide adequate standards and resources for vehicle trials because they have an inadequate representation of vehicle technology and vehicle trials, thus they are unable to determine what is necessary for safe trials	**RSR-6.2:** The company management must have an adequate model of the vehicle technologies and the vehicle trials
Department authorizing vehicle trial	**UCA-7:** The department authorizing the vehicle trial authorizes a non-compliant or/and unsafe vehicle trial	**SR-7:** The department authorizing the vehicle trial must not authorize a non-compliant or/and unsafe vehicle trial	Inadequate mental model: Safety and compliance of the vehicle trial	The department authorizing the vehicle trial authorizes a non-compliant or/and unsafe vehicle trial because they are not aware that the trial is non-compliant or/and unsafe	**RSR-7.1:** The department authorizing the vehicle trial must have an adequate representation of the compliance and safety of the trial
					RSR-7.2: The trial manager, company experts and department providing the prototype, must provide the department authorizing the vehicle trial with adequate feedback on the compliance and safety of the trial
Company Experts	**UCA-8:** The company experts provide inadequate recommendation to ensure the compliance and safety of the trial	**SR-8:** The company experts must provide adequate recommendation to ensure the compliance and safety of the trial	Inadequate mental model: importance of providing recommendations	The company experts provide inadequate recommendations because they do not consider it as a priority in their job functions, thus they do not respond to the request	**RSR-8.1:** The company management must explicitly incorporate the function of providing recommendations for vehicle trials into the company expert's job functions
			Inadequate mental model: Vehicle technology and vehicle trial	The company experts provide inadequate recommendations because they have an inadequate model of the vehicle technology and the vehicle trial	**RSR-8.2:** The company experts must have an adequate model of the vehicle technology and the vehicle trial
					RSR-8.3: The trial manager must provide adequate feedback on the vehicle technology and vehicle trial in the recommendation request
			Inadequate mental model: applicable frameworks	The company experts provide inadequate recommendations because they are unaware of all the frameworks that are applicable to vehicle trials	**RSR-8.4:** The company experts must have an adequate model of the frameworks applicable to vehicle trials

Controller	STPA step 1	Safety Requirement	STPA step 2		
	Unsafe control Action		Control flaw	Scenario	Refined Safety Requirement
Department providing the prototype	**UCA-9:** The department providing the prototype gives consent to use the prototype in an unsafe trial	**SR-9:** The department providing the prototype must not give consent to use the prototype in an unsafe trial	Inadequate mental model: Safety of vehicle technology	The department providing the prototype for the trial grants consent to use an unsafe prototype because they believe that the prototype has an adequate level maturity and safety	**RSR-9.1:** The department providing the prototype must have an adequate model of the prototype's level of maturity and safety
					RSR-9.2: The service providers that participate in the development of the trial must provide adequate feedback on the prototype's level of maturity and safety
			Inadequate mental model: Safety of the vehicle trial	The department providing the prototype for the trial grants consent to use a prototype that is unsafe under the trial conditions because they think that the prototype can have a safe operation during the trial	**RSR-9.3:** The department providing the prototype must have an adequate model of the vehicle trial
					RSR-9.4: The trial manager must provide adequate feedback on the vehicle trial to the department providing the prototype
Trial manager	**UCA-10:** The trial manager defines trial objectives and conditions that contribute to an unsafe trial	**SR-10:** The trial manager must not define trial objectives and conditions that contribute to a safe trial	Inadequate mental model: Vehicle trial hazards	The trial manger defines trial objectives and conditions that contribute to an unsafe trial because s/he is not aware of the vehicle trial hazards	**RSR-10.1:** The trial manager must have an adequate model of the vehicle trial hazards
			Inadequate mental model: Vehicle technology	The trial manger defines trial objectives and conditions that contribute to an unsafe trial because s/he has an inadequate model of the vehicle technology	**RSR-10.2:** The trial manager must have an adequate model of the vehicle technology, its operational design domain and limits
			Inadequate mental model: How to define adequate objectives and conditions	The trial manger defines trial objectives and conditions that contribute to an unsafe trial because s/he thinks that the trial objectives and conditions are adequate	**RSR-10.3:** The trial manger must define clear, measureable, and feasible objectives and conditions for the trial and disseminate them to all the stakeholders concerned by the trial
	UCA-11: The trial manger does not adequately coordinate the trial	**SR-11:** The trial manger must adequately coordinate the trial	Inadequate mental model: Assign responsibilities and resources	The trial manager does not adequately coordinate the trial because s/he does not consider it important to properly assign responsibilities and resources to all the stakeholders of the trial	**RSR-11.1:** The trial manager must assign responsibilities and resources to all the stakeholders of the trial
			Inadequate mental model: Coordination procedures	The trial manager does not adequately coordinate the trial because s/he does know how to properly coordinate a trial or because s/he does not communicate with all the actors of the transport system that are concerned by the trial	**RSR-11.2:** The company management must provide a company standard and training on how to coordinate a trial
					RSR-11.3: The company management must communicate with all stakeholders e.g. local government, the public, emergency services, etc.

Controller	STPA step 1	Safety Requirement	STPA step 2		
	Unsafe control Action		Control flaw	Scenario	Refined Safety Requirement
Trial manager	**UCA-12:** The trial manger inadequately assesses trial compliance and safety	**SR-12:** The trial manger must adequately assess trial compliance and safety	Inadequate mental model: Trial compliance and safety requirements	The trial manager does not adequately assess trial and safety and compliance because s/he does not know what are the requirements that make a trial compliant and safe	**RSR-12.1:** The trial manger must have an adequate model of the requirements that make a trial compliant and safe
					RSR-12.2: The stakeholders that support the assessment of the compliance and safety of the trial must provide the trial manager with adequate feedback
Trial executor(s)	**UCA-13:** The trial executor(s) defines a protocol that contributes to unsafe or/and non-compliant vehicle trials	**SR-13:** The trial executor(s) must not define a protocol that contributes to unsafe or/and non-compliant vehicle trials	Inadequate mental model: Safety and compliance of vehicle trial	The trial executor(s) defines a protocol that contributes to unsafe vehicle trials because s/he is not are not aware that the trial protocol is unsafe or/and non-compliant	**RSR-13.1:** The trial executor must have an adequate model of the safety and compliance of the protocol
					RSR-13.2: The trial executor must conduct a risk analysis of the trial and establish risk management strategies
					RSR-13.3: The trial executor must verify the compliance of the trial
	UCA-14: The trial executor(s) defines inadequate data recording and processing specifications	**SR-14:** The trial executor(s) must define adequate data recording and processing specifications	Inadequate mental model: data needs for the trial objectives and liability matters	The trial executor(s) defines inadequate data recording specifications because s/he is not aware of the data that is needed to achieve the objectives of the trial or/and to be prepares for liability matters in case of an incident	**RSR-14.1:** The trial executor must have an adequate model of the data needed to achieve the trial objectives and to clear liability matters
					RSR-14.2: The trial executor must have an adequate model of the legislation regarding personal data
			Inadequate mental model: data protection framework	The trial executor(s) defines inadequate data recording specifications because s/he has an inadequate model of the personal data protection framework	**RSR-14.3:** A company expert must validate that data recording and processing specifications are compliant with the data protection framework
	UCA-15: The trial executor(s) collaborates in an inadequate fashion with company teams and service providers	**SR-15:** The trial executor(s) must collaborate in an adequate fashion with company teams and service providers	Inadequate mental model: Safety of the requirements and specifications	The trial executor(s) collaborates in an inadequate fashion with the company teams and service providers because s/he defines inconsistent or unsafe requirements and specifications to the company teams and service providers	**RSR-15.1:** The trial executor(s) must provide consistent and safe requirements to the company teams and service providers
					RSR-15.2: There must be adequate communication between the trial executor(s), the company teams and service providers to discuss problems and solutions for the implementation of the requirements and specifications
	UCA-16: The trial executor(s) prepare the trial in an inadequate fashion	**SR-16:** The trial executor(s) must prepare the trial in an adequate fashion	Inadequate mental model: Trial needs	The trial prepares the trial in an inadequate fashion because s/he is not aware of all the trial needs (logistics, participant recruitment, etc.)	**RSR-16.1:** The trial executor must have an adequate model of the trial needs

Controller	STPA step 1	Safety Requirement	STPA step 2		
	Unsafe control Action		Control flaw	Scenario	Refined Safety Requirement
Company teams and service providers	**UCA-17:** The company teams and service providers inadequately implement the requirements and specifications given by the trial executors	**SR-17:** The company teams and service providers must adequately implement the requirements and specifications given by the trial executors	Inadequate mental model: Trial requirements and specifications	The company teams and service providers inadequately implement requirements because they have an inadequate model of the requirements, specifications and the actions to implement them	**RSR-17.1:** The company teams and service providers must have an adequate model of the trial requirements
					RSR-17.2: The trial executor must provide company teams and service providers with correct, complete, and safe trial requirements and specifications
					RSR-17.3: The company teams and service providers must verify the adequate implementation of the trial requirements and specifications
			Inadequate mental model: Resources	The company teams and service providers inadequately implement requirements because they believe that they have the resources to adequately implement the trial requirements, when in reality.	**RSR-17.4:** The company management must provide resources to the company teams that enable an adequate implementation of the trial requirements
					RSR-17.5: The trial manger must verify that the service providers have the resources to adequately implement trial requirements

Appendix C: Results of the STPA on the vehicle trial operation (chapter 4)

Table 48 – STPA on the vehicle trial involving a highway pilot system

Controller	Control Action	Unsafe control action	High-level safety requirement	Control flaw	Scenario	Refined Safety Requirement
Trial staff	Follow instructions	**UCA-1:** The trial staff follows instructions on trial logistics, secure trial site, vehicle and people safety, in an inadequate fashion	**SR-1:** The trial staff must adequately follow instructions to manage logistics, to secure the trial and to ensure the safety of people involved in the trial	Inadequate feedback: Trial instructions	The trial staff inadequately follows instructions because the trial instructions are incorrect, incomplete, unclear or difficult to understand	**RSR-1.1:** The trial design team must provide the trial staff with adequate, correct, complete and understandable trial instructions.
				Inadequate model: How to follow trial instructions	The trial staff inadequately follows instructions because they are unaware of how to follow the instructions	**RSR-1.2** The trial staff must have an adequate model of how to follow the trial instructions
				Inadequate model: Vehicle technology	The trial staff inadequately follows because they have an inadequate model of the vehicle technology on which they verify sensor performance and provide technical assistance	**RSR-1.3:** The trial staff must have an adequate model of the vehicle technology
Trial experimenter	Follow protocol	**UCA-2:** The trial experimenter follows the protocol in an inadequate fashion	**SR-2:** The trial experimenter must adequately follow the trial protocol	Inadequate feedback: Trial protocol	The trial experimenter inadequately follows the protocol because the protocol contains incorrect, incomplete, unclear instructions	**RSR-2.1:** The trial design team must provide the trial experimenter with an adequate, correct, complete and understandable trial protocol.
				Inadequate model: How to follow protocol	The trial experimenter inadequately follows the protocol because s/he does not know how to follow the instructions of the protocol	**RSR -2.2** The trial experimenter must have an adequate model of how to follow the trial protocol
	Provide instructions	**UCA-3:** The trial experimenter provides instructions to the participant (how to operate the vehicle technology, what to do, safety instructions, etc.) in an inadequate fashion	**SR-3:** The trial experimenter must adequately provide instructions to the participant (how to operate the vehicle technology, what to do, safety instructions, etc.)	Inadequate feedback: Instructions	The trial experimenter inadequately provides instructions to the driver participant because the trial experimenter receives inadequate feedback on instructions	**RSR-3.1:** The design team must provide the trial experimenter with adequate feedback on the driver participant instructions
				Inadequate model: How to provide instructions	The trial experimenter inadequately provides instructions to the driver participant because the trial experimenter does not know how to provide instructions	**RSR-3.2:** The trial experimenter must receive training on how to provide instructions to the driver participant
				Ineffective control action: Instructions	The trial experimenter inadequately provides instructions to the driver participant because the driver participant does not understand the instructions	**RSR-3.3:** The trial experimenter must perform pre-tests to verify that s/he is able to adequately provide instructions and that the driver participant can understand them

Controller	Control Action	Unsafe control action	High-level safety requirement	Control flaw	Scenario	Refined Safety Requirement
Trial experimenter	Record data	**UCA-4:** The trial experimenter does not record data during the vehicle trial	**SR-4:** The trial experimenter must record data during the vehicle trial	Inadequate feedback: trial instructions and trial training	The trial experimenter does not record data during the vehicle trial because the trial instructions or/and trial training contain inadequate feedback regarding how to record trial data	**RSR-4.1:** The trial experimenter must receive adequate feedback on how to record data in the trial instructions and trial training
				Missing control action: Record data	The trial experimenter does not record data during the vehicle trial because the trial experimenter forgets to launch the data recorder	**RSR-4.2:** The trial protocol must include a step to launch the data recorder
				Inadequate sensor operation: Record data	The trial experimenter does not record data during the vehicle trial because the trial instructions launches the data recorder but no data is recorded	**RSR-4.3:** The trial experimenter must verify that the data recorder is recording data before the start of the driving phase of every trial
	Interact with the participant	**UCA-5:** The trial experimenter interacts with the driver participant and that causes the driver to put the vehicle in an unsafe situation	**SR-5:** The trial experimenter must not cause the driver participant to put the vehicle in an unsafe situation when the trial experimenter interacts with the driver participant	Inadequate model: Stress and distraction	The trial experimenter interacts with the driver participant and causes the driver to put the vehicle in an unsafe situation because the experimenter believes that s/he is not generating stress or distraction to the driver	**RSR-5.1:** The trial experimenter protocol must limit the interactions with the driver participant that generate stress and distraction
Driver participant	Control the vehicle	**UCA-6:** The driver provides inadequate control of the vehicle or/and violates traffic rules during manual driving mode	**SR-6:** The driver must provide adequate control of the vehicle and comply with traffic rules during manual driving mode	Inadequate feedback: State of the driving mode	The driver provides inadequate control of the vehicle or/and violates traffic rules during manual driving because s/he does not perceive that the vehicle is on manual mode	**RSR-6.1:** The HMI must have an adequate and robust operation that enables the driver participant to determine the state of the driving mode
				Inadequate Model: State of the driving mode	The driver provides inadequate control of the vehicle or/and violates traffic rules during manual driving because s/he has an inadequate model of the state of the driving mode (s/he thinks the vehicle is on AD mode)	**RSR-6.2:** The driver must have an adequate model of the current driving mode and the HMI interfaces for the two driving modes
						RSR-6.3: The trial design team must define actions for the trial experimenter and the trial supervisor in case the driver experiences mode confusion
	Release control of the vehicle	**UCA-7:** The driver does not release the control of the vehicle after the automated driving system engagement	**SR-7:** The driver must not release the control of the vehicle after the automated driving system engagement	Inadequate feedback: HMI	The driver does not release the control of the vehicle after the automated driving system engagement because s/he does not perceive that the ADS is engaged	**RSR-7.1:** The HMI must provide adequate feedback on ADS engagement
				Inadequate model: Transition to AD mode	The driver does not release the control of the vehicle after the automated driving system engagement because s/he does not understand the request or/and how to respond to the request	**RSR-7.2:** The HMI must provide adequate information on the takeover request
						RSR-7.3: The driver participant training must cover the HMI information and sequences to transition to AD mode
						RSR-7.4: The trial design team must define actions for the trial experimenter and the trial supervisor in case the driver does not release control after ADS engagement

Controller	Control Action	Unsafe control action	High-level safety requirement	Control flaw	Scenario	Refined Safety Requirement
Driver participant	Release control of the vehicle	**UCA-8:** The driver releases the control of the vehicle before the automated driving system is engaged	**SR-8:** The driver must not release control of the vehicle before the automated driving system is engaged	Inadequate model: AD engagement	The driver releases the control of the vehicle before the automated driving system is engaged because s/he does not understand the feedback provided by the HMI and believes that the ADS is engaged	**RSR-8.1:** The HMI must indicate when the driver can release control
				Inadequate model: Transition to AD mode	The driver releases the control of the vehicle before the automated driving system is engaged because s/he does knows that the ADS is not yet engaged, but is unaware of the action sequences to transition to AD mode, and believes that s/he has to release control of the vehicle	**RSR-8.2:** The driver participant training must cover the ADS engagement procedure, notably when to release control of the vehicle
						RSR-8.3: The trial design team must define actions for the trial experimenter and the trial supervisor in case the driver releases control before the engagement of the ADS
	Respond to a takeover request	**UCA-9:** The driver does not respond to the takeover request	**SR-9:** The driver should respond to the takeover request	Missing feedback: Takeover request	The driver participant does not respond to the takeover request because s/he does not perceive the request	**RSR-9.1:** The HMI must provide salient and intuitive feedback regarding the takeover request
				Inadequate mental model: how to respond to the takeover request	The driver participant does not respond to the takeover request because s/he does not know how to respond to the takeover request	**RSR-9.2:** The driver participant training must cover how to respond to the takeover request
						RSR-9.3: The trial design team must define actions for the trial experimenter and the trial supervisor in case the driver does not respond to the takeover request
		UCA-10: The driver responds to the takeover request when s/he has not regained situation awareness	**SR-10:** The driver must not put the vehicle in an unsafe situation when s/he has not regained situation awareness	Inadequate feedback: HMI feedback on the takeover request	The driver participant responds to the takeover request and puts the vehicle in an unsafe situation because s/he does not understand the feedback provided by the HMI	**RSR-10.1:** The HMI must provide clear and understandable feedback regarding the takeover request
				Inadequate mental model: Respond to takeover request before regaining SA	The driver participant responds to the takeover request and puts the vehicle in an unsafe situation because s/he has an incorrect representation of the takeover procedure, thinks that s/he must immediately respond to the request even if s/he has not regained situation awareness	**RSR-10.2:** The HMI must suggest that the driver needs to regain situation awareness before responding to the request
						RSR-10.3: The driver participant training must cover the importance of regaining situation awareness before responding to the takeover request
						RSR-10.4: The trial design team must define actions for the trial experimenter and the trial supervisor in case the driver has not regained situation awareness

Controller	Control Action	Unsafe control action	High-level safety requirement	Control flaw	Scenario	Refined Safety Requirement
Driver participant	Respond to a takeover request	**UCA-10:** The driver responds to the takeover request when s/he has not regained situation awareness	**SR-10:** The driver must not put the vehicle in an unsafe situation when s/he has not regained situation awareness	Inadequate model: traffic environment	The driver participant inadequately responds to the takeover request because s/he has an inadequate representation of the traffic environment when s/he responds to the request	**RSR-10.5:** The trial design team must define actions for the trial experimenter and the trial supervisor in case the driver puts the vehicle in an unsafe situation when s/he responds to the takeover request
Trial supervisor (co-driver)	Intervene	**UCA-11:** trial supervisor does not intervene when safety is threatened	**SR-11:** The trial supervisor must intervene when safety is threatened	Inadequate model: Unsafe situations	The trial supervisor does not intervene when safety is threatened because s/he is not aware that safety is threatened	**RSR-11.1:** The trial supervisor must have an adequate model of the situations in which safety is threatened
		UCA-12: The trial supervisor intervenes and puts the vehicle in an unsafe situation	**SR-12:** The trial supervisor must not put the vehicle in an unsafe situation when s/he intervenes	Inadequate feedback: Traffic environment	The trial supervisor puts the vehicle in an unsafe situation when s/he intervenes because s/he collides with another vehicle on the adjacent lane due to missing perception of the other road user	**RSR-12.1:** The trial supervisor must perceive the traffic environment when s/he intervenes
				Inadequate model: Intervention needed and how to intervene	The trial supervisor puts the vehicle in an unsafe situation when s/he intervenes because s/he has an inadequate model of how to intervene in case of an emergency	**RSR-12.2:** The trial supervisor must have an adequate model to determine which situations need intervention and to adequately intervene (operate the override actuators, emergency switch and stop button)
				Inappropriate control action: intervention	The trial supervisor puts the vehicle in an unsafe situation when s/he intervenes because s/he provides inappropriate control actions when s/he intervenes	**RSR-12.3:** The trial supervisor must be trained to learn how to execute appropriate control actions when s/he intervenes
				Inadequate actuator operation	The trial supervisor puts the vehicle in an unsafe situation when s/he intervenes because the actuators that enable the intervention (double commands, miniwheel, emergency switch and stop button) have an inadequate operation	**RSR-12.4:** The actuators that enable the trial supervisor 's intervention must have an adequate operation
Automation	Send AD is available	**UCA-13:** Automation sends “AD is available” when AD is not available	**SR-13:** Automation must not send “AD is available” when AD is not available	Inadequate feedback: AD availability data	Automation sends “AD is available” when AD is not available because the feedback provided by vehicle sensors or external information, indicate that AD availability conditions are met when they are not	**RSR-13.1:** The trial design team must verify that the perception system of the automated driving system and external information, do not provide measures on AD availability conditions indicating that they are met, when they are not

Controller	Control Action	Unsafe control action	High-level safety requirement	Control flaw	Scenario	Refined Safety Requirement
Automation	Send AD is available	**UCA-13:** Automation sends "AD is available" when AD is not available	**SR-13:** Automation must not send "AD is available" when AD is not available	Inadequate model: AD availability	Automation sends "AD is available" when AD is not available because automation believes that AD is available when it is not	**RSR-13.2:** The trial design team must verify that the automated driving system does not send "AD is available" when it is not available
					Automation sends "AD is available" when AD is not available because automation believes that AD is available when it is not	**RSR-13.3:** The trial design team must conduct pre-trials in which the vehicle is operated on AD mode on the trial route to detect the segments in which AD is potentially available, and limit AD availability proposals
	Control the vehicle	**UCA-14:** Automation provides control of the vehicle during manual driving	**SR-14:** Automation must not provide control of the vehicle during manual driving	Inadequate control algorithm: provide control	Automation provides control of the vehicle during manual driving because automation's control algorithm generates actions to control the vehicle when the vehicle is on manual mode	**RSR-14.1:** The trial design team must verify that automation's control algorithm does not generate actions to control the vehicle when the vehicle is on manual mode
						RSR-14.2: The trial supervisor must intervene and override automation when automation provides control of the vehicle during manual driving mode
		UCA-15: Automation provides inadequate control of the vehicle during automated driving	**SR-15:** Automation must provide adequate control of the vehicle during automated driving	Inadequate feedback: Driving environment	Automation provides inadequate control of the vehicle during automated driving because the feedback provided by vehicle sensors on the driving environment is inadequate	**RSR-15.1:** The trial design team must verify that the perception system of the automated driving system provide adequate feedback on the driving environment
				Inadequate Model: Driving environment	Automation provides inadequate control of the vehicle during automated driving because automation has an inadequate representation of the driving environment	**RSR-15.2:** The trial design team must verify that automation has an adequate representation of the driving environment
				Inadequate control actions: control of the vehicle	Automation provides inadequate control of the vehicle during automated driving because automation has an inadequate representation of the driving environment	**RSR-15.3:** The trial design team must conduct pre-trials in which the vehicle is operated on AD mode on the trial route to validate that automation executes adequate actions to control the vehicle
						RSR-15.4: The trial supervisor must intervene and override automation when automation provides inadequate control of the vehicle

Controller	Control Action	Unsafe control action	High-level safety requirement	Control flaw	Scenario	Refined Safety Requirement
	Send takeover request	**UCA-16:** Automation does not send a takeover request when it reaches the limits its operational design domain	**SR-16:** Automation must send a takeover request when it reaches the limits of its operational design domain	Inadequate feedback: operational design domain	Automation does not send a takeover request when it reaches the limits of its operational design domain because the feedback provided by vehicle sensors or external information, does not indicate that automation has reached its operational design domain	**RSR-16.1:** The trial design team must verify that the perception system of the automated driving system and external information, enable automation to determine when it reaches its operational design domain
				Inadequate model: operational design domain	Automation does not send a takeover request when it reaches the limits of its operational design domain because automation has an inadequate representation of its design domain, automation s unaware that it has reached its operational design domain	**RSR-16.2:** The trial design team must verify that the automated driving system does not send "AD is available" when it is not available
						RSR-16.3: The trial design team must conduct pre-trials in which the vehicle is operated on AD mode on the trial route to validate that automation can determine when it is outside its operational design domain and send a takeover request
	MRM	**UCA-17:** Automation does not provide an MRM when the driver does not respond to the takeover request (end modes types 1 and 2) or when there is a performance-relevant failure (end mode type 3)	**SR-17:** Automation must provide an MRM when the driver does not respond to the takeover request (end modes types 1 and 2) or when there is a performance-relevant failure (end mode type 3)	Missing control action: MRM	Automation does not provide an MRM when the driver does not respond to the takeover request (end modes types 1 and 2) or when there is a performance-relevant failure (end mode type 3) because the control actions necessary to perform an MRM are not provided by automation's control algorithm	**RSR-17.1:** The trial design team must conduct pre-trials in which the vehicle is operated on AD mode on the trial route to validate that automation's control algorithm provides MRM
						RSR-17.2: The trial supervisor must intervene when automation does not provide a MRM
		UCA-18: Automation provides an MRM and puts the vehicle in an unsafe situation	**SR-18:** Automation must not put the vehicle in an unsafe situation when it provides an MRM	Inappropriate control action: MRM	Automation provides an MRM and puts the vehicle in an unsafe situation because automation executes inappropriate control actions to execute the MRM	**RSR-18.1:** The trial design team must conduct pre-trials in which the vehicle is operated on AD mode on the trial route to validate that automation does not put the vehicle in unsafe situations when automation executes an MRM
						RSR-18.2: The trial supervisor must intervene when automation puts the vehicle in an unsafe situation while automation executes an MRM

Résumé

Les constructeurs automobiles fabriquant des systèmes de conduite automatisée ont besoin d'aborder les conséquences que ces systèmes peuvent avoir sur la sécurité routière. Notamment pour l'évaluation des gains de sécurité, la sécurisation des essais et l'analyse des accidents impliquant le véhicule autonome. Cependant, le cadre conceptuel actuel utilisé dans la sécurité routière peut ne pas être adapté pour l'analyse des changements et des nouvelles interactions introduits par l'automatisation du véhicule à travers toutes les échelles du système sociotechnique de transport routier.

Le but de la thèse est d'appliquer une approche systémique fondée sur STAMP afin étudier les gains attendus du véhicule autonome en termes de sécurité routière, sécuriser les expérimentations et analyser les accidents impliquant ce type de véhicule, à travers toutes les échelles du système sociotechnique de transport routier.

Afin de contribuer au calcul des gains du véhicule autonome sur la sécurité routière, la population cible d'un « highway pilot system» a été définie et des questions issue d'une analyse STPA (analyse des risques issue de STAMP) aidant à l'évaluation de l'efficacité du système ont été élaborées.

Un cadre de sécurisation des expérimentations couvrant tous les niveaux du système a été mis en place au moyen d'une analyse STPA à deux échelles.

Enfin, une méthode d'analyse des accidents impliquant un conducteur automatisé a été créé en intégrant des éléments issus de méthodes d'analyses des accidents de la route existantes et des éléments explicatifs développés spécialement à la méthode CAST (méthode d'analyse des accidents fondée sur STAMP). L'accident impliquant une Tesla en mai 2016 est le cas d'étude de cette nouvelle méthode, CASCAD.

En conclusion, ces trois applications ont montré tout le potentiel d'une approche systémique fondée sur STAMP pour offrir un cadre conceptuel adapté à l'évaluation des conséquences sur la sécurité routière de la conduite automatisée

Mots Clés

Véhicule autonome, Sécurité Routière, Gains de sécurité, Sécurisation des essais, Analyse des accidents de la route, STAMP

Abstract

As automakers develop automated driving systems, they must address the implications of such systems on road safety. Notably for the safety benefit assessment, trial safety and accident analysis. However, the existing conceptual framework in road safety may not be adapted to analyze the changes and new interactions introduced by vehicle automation at all the levels of the road transport system.

The main objective of this thesis is to apply a systems theoretic approach based on STAMP to examine the safety benefit assessment, trial safety and accident analysis of automated driving across all the levels of the road transport sociotechnical system.

This research first contributes to the safety benefit assessment by estimating the target population of a highway pilot system and by generating questions derived from an STPA analysis (hazard analysis based on STAMP) to facilitate the evaluation of the influence of the highway pilot system on road safety.

Next, this work establishes a framework to ensure trial safety across the macroscopic and microscopic levels of the vehicle trial system by structuring the outputs of two STPA analyses.

Finally, this thesis integrates elements from existing crash analysis methods and newly developed guidance elements into CAST (an accident analysis method based on STAMP) to develop a new method for the accident analysis of crashes involving automated driving called CASCAD. The application of CASCAD is illustrated using the available information of the Tesla crash on May 2016.

The three applications of this research show the potential of a STAMP-based approach to provide a suitable conceptual framework for the analysis of the implications of road safety on automated driving..

Keywords

Automated driving, Road Safety, Safety benefit, Trial Safety, Crash analysis, STAMP