Deliverable **D3.1** /

From Research Questions to Logging Requirements

Version: 1.0 Final
Dissemination level: PU
Lead contractor: University of Leeds
Due date: 31.03.2018
Version date: 24.04.2018

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 723051.

Document information

AUTHORS

Daryl Hibberd – University of Leeds
Tyron Louw – University of Leeds
Elina Aittoniemi – VTT
Afsaneh Bjorvatn – SNF
Rino Brouwer – TNO
Mandy Dotzauer – DLR
Felix Fahrenkrog - BMW
Satu Innamaa – VTT
Salla Kuisma – VTT
Natasha Merat – University of Leeds
Barbara Metz – WIVW
Natalia Neila – JLR
Karl Pedersen – SNF
Merja Penttinen – VTT
Pablo Puente Guillen – TME
Christian Rösener – ika
Anne Silla – VTT
Thomas Streubel – SAFER
Fabio Tango – CRF
Benedikt van den Boom – FIA
Hendrik Weber – ika
Johanna Woerle – WIVW
Alexander Zerbe – BAST

Coordinator

Aria Etemad
Volkswagen Group Research
Hermann-Münch-Str. 1
38440 Wolfsburg
Germany

Phone: +49-5361-9-13654
Email: aria.etemad@volkswagen.de

Project funding

Horizon 2020
ART-02-2016 – Automation pilots for passenger cars
Contract number 723051
www.L3Pilot.eu

Legal Disclaimer

Revision and history chart

Version	Date	Comment
0.1	24/11/2017	Draft table of contents
0.2	27/11/2017	Draft table of contents edits
0.3	01/12/2017	Draft tables of contents edits
0.4	05/12/2017	Structure updated and partner contributions assigned
0.5	19/12/2017	Contributions added and new partners allocated to sections
0.6	08/01/2018	Contributions added. Edits of existing sections. Some addition of partners to sections.
0.7	15/01/2018	Contribution added. Urban chauffeur literature review
0.8	09/02/2018	Contributions added. Main edits Chapter 1.5
0.9	07/03/2018	Contributions added. RQ table added. Edits before submission of draft version.
1.0	11/04/2018	Edits made in response to preliminary internal review
1.1	20/04/2018	Edits made in response to formal internal review
2.0	24/04/18	Final proof-read, reference check and formatting – Final draft

Table of contents

List of figures

List of tables

Executive summary

L3Pilot is a large-scale, Europe-wide, real-world pilot study of SAE Level 3 automated driving functions. The functions to be tested include a Motorway Chauffeur, Urban Chauffeur, and Parking Chauffeur, all designed to enable the driver to move safely and efficiently through selected road environments. This deliverable reports the results of applying the 'FESTA V' methodology to the development of a set of Research Questions that will be used to direct the study of SAE Level 3 (and Level 4) automated driving.

The main result reported here is a list of Research Questions (RQs) with associated hypotheses, performance indicators and required vehicle-based data logging needs, agreed by partners, following a series of iterations[1]. These RQs are tied to a specific Evaluation Area and Impact Area which were defined at the outset of the L3Pilot project, and are motivated by the FESTA approach and existing literature in the field. The RQ-list is intended as a 'living document', to be reviewed and revised as more detailed information becomes available about the ADFs for test (reported in D4.1), and further work is completed in the matching of Experimental Methods and Evaluation Methods to Research Questions. The document is delivered with an accompanying Table showing the selected Research Questions, with related hypotheses and performance indicators in additional detail.

The output from this Deliverable feeds into the work conducted by WP3.4 and WP3.5, where experimental procedures and evaluation methods are mapped on to this Research Questions list, and the feasibility of each RQ is assessed through interaction with partners involved in the preparation and delivery of the pilot studies, and the evaluation activities (SP7).

This deliverable also informs the pre-piloting stage of the ADFs in SP4, such that the experimental studies can be designed and set-up with the Research Questions and hypotheses in mind. The definition of Performance Indicators required for the evaluation of the ADFs supports the production of piloting tools and testing of the data processing and analysis chain (SP5), before full-scale, on-road, testing begins (SP6).

This work will also provide contributions to the Code of Practice for the Development of ADFs (SP2); specifically the amendments to the FESTA methodology to accommodate testing of SAE Level 3 and Level 4 ADFs.

The final version of the RQs that will be addressed in the L3Pilot project will be reported in D3.4 Final Evaluation Plan.

---

[1] To allow for timely delivery of the results, final RQs were agreed by 7th March 2018.

Glossary

The Glossary provides a list of key terms in this deliverable and their definitions based on previous work in the field. A more extensive glossary for the whole project will be developed as a standalone document. The Glossary definitions below were correct as of 07 March 2018.

Term	Meaning
Automated Driving Function	Activity or purpose of a vehicle to enable automated driving.
Automated Driving System	A combination of hardware and software required to realise an ADF.
Assist	Automated driving function operating at SAE L2.
Baseline	Set of data to which the performance and the effects of the technology under study are compared.
CAN Signals	Data from the vehicle internal data communication, collected from the vehicles.
Chauffeur	Automated driving function operating at SAE L3.
Derived Measures	A single measure calculated from a direct measure (e.g. by applying mathematical or statistical operations) or a combination of one or more direct or derived measures (FOT-Net Data, 2016, pp. 55-56).
Direct Measures	A measure logged directly from a sensor, without further manipulations except linear transformations (e.g. m/s to kph) before saving the data to the log file (FOT-Net Data, 2016, p. 55).
Driving Scenarios	The abstraction and the general description of a driving situation without any specification of the parameters of the driving situation, thus, it summarises a cluster of homogenous driving situations. Driving scenarios are typically short in time (t < 30 s) and only a few vehicles are involved. An example is lane change to the left lane (AdaptIVe).
Driving Situation	A specific driving manoeuvre (e.g. a specific lane change with defined parameters). Thus the driving situation describes in detail a driving scenario e.g. a lane change at 60.8 distance of 10 m behind the host vehicle in the adjacent lane and with a velocity of 65 km/h.
Events	Events are either single time-points or segments of time in time-series data for which one or several criteria are fulfilled. An event can be short (e.g. crash) or long, such as, for example, the start of an evasive manoeuvre, car following, overtaking, or speeding.
Hypothesis	A specific statement linking a cause to an effect and based on a mechanism linking the two. It is applied to one or more functions and can be tested with statistical means by analysing specific performance indicators in specific scenarios. A hypothesis is expected to predict the direction of the expected change (FOT-Net Data, 2016, p. 48).
Operational Design Domain	The specific conditions under which a given driving automation system or feature thereof is designed to function, including, but not limited to, driving modes (SAE, 2016).

Term	Meaning
Performance Indicator	Quantitative or qualitative indicator[s], derived from one or several measures, agreed on beforehand, expressed as a percentage, index, rate or other value, which are monitored at regular or irregular intervals and can be compared to one or more criteria. (Mäkinen et al., 2011, p. 45). In some cases, these will be the same as a derived measure, in other cases, further processes are required to generate a PI.
Pilot	Automated driving function operating at SAE L4.
Pilot Test	Field test of applications and functions not as mature as in FOTs. The methodology for testing, however, may be in principal the same. The test is used to decide how and whether to launch a full-scale project.
Raw Data	Data that has been recorded in instrumented vehicles (CAN data, video, GPS logs etc.). This data is by nature heterogeneous; different vehicles will produce different datasets. These datasets are thus not immediately useful for comparison.
Research Question	A general question to be answered by compiling and testing related specific hypotheses (FOT-Net Data, 2016, p. 39).
SAE L0 - No Automation	The full-time performance by the human driver of all aspects of the dynamic driving task, even when enhanced by warning or intervention systems (SAE Levels of Driving Automation, J3016 SEP2016).
SAE L1 - Driver Assistance	The driving mode-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the human driver performs all remaining aspects of the dynamic driving task (SAE, 2016).
SAE L2 - Partial Automation	The driving mode-specific execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the human driver performs all remaining aspects of the dynamic driving task (SAE, 2016).
SAE L3 - Conditional Automation	The driving mode-specific performance by an Automated Driving System of all aspects of the dynamic driving task with the expectation that the human driver will respond appropriately to a request to intervene (SAE, 2016).
SAE L4 - High Automation	The driving mode-specific performance by an Automated Driving System of all aspects of the dynamic driving task, even if a human driver does not respond appropriately to a request to intervene (SAE, 2016).
SAE L5 - Full Automation	The full-time performance by an Automated Driving System of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver (SAE, 2016).
Self-Reported Measures	Subjective data reported or observed through various tools (FOT-Net Data, 2016, p. 56).
Situational Variable	An aspect of the surroundings made up of distinguishable levels. At any point in time at least one of these levels must be valid (FOT-Net Data, 2016, p. 57).
Traffic Scenario	A traffic scenario describes a larger traffic context, which includes different (not pre-defined) driving scenarios. Typically in a traffic scenario a large number of vehicles is analysed over a longer time period. An example of a traffic scenarios could be on a three-lane motorway section with ten highway entrances and exits, a speed limit of 130kph for a period of one hour. (AdaptIVe)
Use Cases	A list of actions or event steps in which a system or its specific function is expected to interact with a user or another system to achieve a goal (FOT-Net

Term	Meaning
	Data, 2016, p. 42). This includes the interaction with users, defined as “anyone who uses the road” (CARTRE, 2017, p. 3).
User	A general term referencing the human role in driving automation (SAE, 2016).

1 Introduction

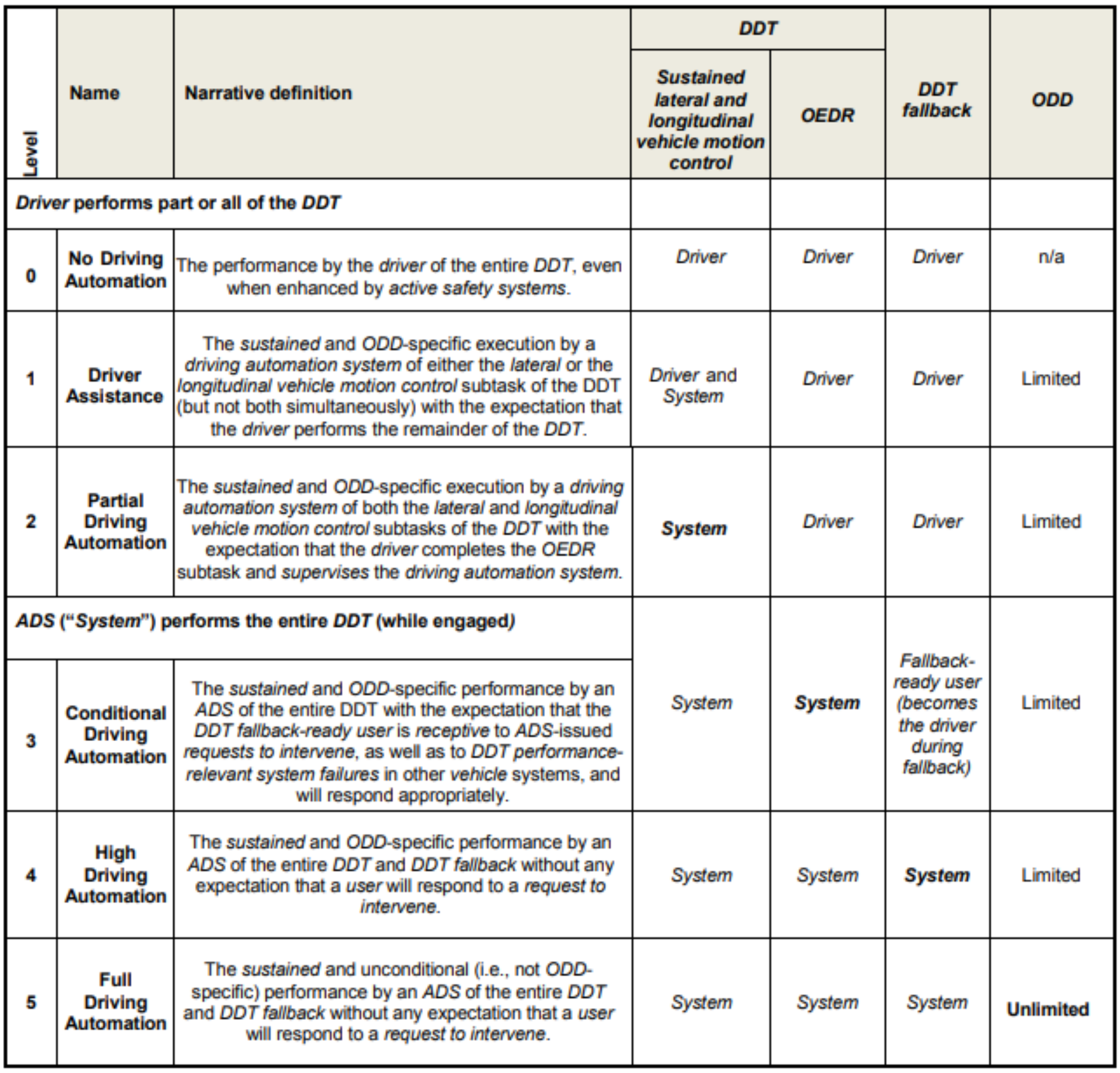
1.1 Background to the L3Pilot Project

Over the years, numerous projects have paved the way for automated driving (AD). Significant progress has been made, but AD is not yet ready for market introduction. However, the technology is rapidly advancing, and today we are at a stage that justifies the large-scale pilot-testing of automated driving. AD technology has matured to a level motivating a final phase of road tests, which can answer key questions, before market introduction.

Automation is not achieved simply by integrating more and better technology. The implementation of automation and deployment of automated vehicles on our roads needs a focus on understanding driver behaviour, willingness to use, and acceptance of automated driving systems (both from the perspective of the driver and society more generally). This user acceptance is one key aspect of the successful deployment of ADFs, in addition to other factors such as understanding the legal challenges and restrictions, which need to be discussed and solved in this context. It is also crucial to investigate the technical feasibility of novel automated driving systems.

The idea of the vehicle being controlled by a series of computers may create fears among the global populous, akin to those in the 1800s, when the motor vehicle was first introduced. A lack of acceptance may hinder the introduction of driver assistance systems with automation despite their obvious benefits for safety and efficiency. In order to overcome public concerns, automated vehicles (AV) need to be designed according to user needs, otherwise they will not be accepted.

L3Pilot is taking the last steps before the introduction of automated cars in daily traffic. The project will undertake large-scale testing and piloting of AD with developed SAE Level 3 (L3) functions (Figure 1.1) exposed to different users including conventional vehicle drivers and Vulnerable Road Users (VRUs), in mixed traffic environments, along different road networks. Level 4 (L4) functions and connected automation will also be assessed in some cases. It should be noted that an important distinction between Level 2 and Level 3 systems is the shift in monitoring responsibility from the human to the AD system (SAE, 2016). With a Level 2 function, the onus is on the driver to constantly monitor the driving task and driving environment. With a Level 3 function, this is performed by the vehicle, with the driver remaining ready to takeover control should the conditions arise where this is necessary. This difference means that there is a considerable change in the technical capabilities of a Level 3 ADF compared to Level 2.



Level	Name	Narrative definition	DDT		DDT fallback	ODD
			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 *DDT*, even when enhanced by *active safety systems*.	*Driver*	*Driver*	*Driver*	n/a
1	**Driver Assistance**	The *sustained* and *ODD*-specific execution by a *driving automation system* of either the *lateral* or the *longitudinal vehicle motion control* subtask of the DDT (but not both simultaneously) with the expectation that the *driver* performs the remainder of the *DDT*.	*Driver* and *System*	*Driver*	*Driver*	Limited
2	**Partial Driving Automation**	The *sustained* and *ODD*-specific execution by a *driving automation system* of both the *lateral* and *longitudinal vehicle motion control* subtasks of the *DDT* with the expectation that the *driver* completes the *OEDR* subtask and *supervises* the *driving automation system*.	***System***	*Driver*	*Driver*	Limited
***ADS* ("*System*") performs the entire *DDT* (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
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
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**

Figure 1.1: SAE Levels of Driving Automation J3016 SEP2016 (Copyright 2016 SAE International).

Extensive on-road testing is vital to ensure sufficient AD function operating performance, to allow an assessment of ongoing user interaction and acceptance of the system. A large and varied sample of drivers need to be involved in this work, to ensure effective piloting, testing and evaluation of ADFs. This work will investigate four ADFs, performing automated driving tasks in three driving environments: motorway, urban and parking environments. In the motorway environment, there will be functions capable of performing either high speed driving or operating in traffic jams, or both. In this project, L3 systems of this type will be termed ‘chauffeurs’, for example a *Motorway Chauffeur*. L4 systems will be termed ‘pilots’, for example a *Motorway Pilot.* However, this distinction does not necessarily reflect the publicly-marketed names of the AD functions.

The data collected in these extensive pilots will support the main aims of the project to:

- Lay the foundation for the design of future, user-accepted, L3 (and L4) systems, to ensure their commercial success. This will be achieved by assessing user reactions, experiences and preferences relating to the AD systems' functionalities.
- Enable non-automotive stakeholders, such as authorities and certification bodies, to prepare measures that will support the uptake of AD, including updated regulations for the certification of vehicle functions with a higher degree of automation, as well as incentives for the user.
- Create unified de-facto standardised methods to ensure further development of AD applications (Code of Practice).
- Perform detailed data analysis to show the performance and effects of ADFs in all relevant conditions, in terms of weather, visibility and traffic volumes.

The consortium addresses four major technical and scientific objectives listed below:

1. Create a standardised Europe-wide piloting environment for automated driving.
2. Coordinate activities across the piloting community to acquire the required data.
3. Pilot, test and evaluate automated driving functions and connected automation.
4. Innovate and promote AD for wider awareness and market introduction.

The L3Pilot consortium brings together stakeholders from the whole value chain, including: OEMs, suppliers, academic institutes, research institutes, infrastructure operators, governmental agencies, the insurance sector and user groups. More than 1000 users will test approximately 100 vehicles across Europe with bases in 11 European countries, including: Austria, Belgium, Finland, France, Germany, Italy, Luxembourg, the Netherlands, Sweden, Spain and the United Kingdom. The project will last for 48 months, and includes 18 months of road tests.

1.2 Objectives of the Methodology sub-project in L3Pilot

The objectives of the *Methodology* sub-project in L3Pilot are to:

- Develop a methodology for the piloting, testing and evaluation of AD systems for achieving reliable results;
- Reconsider the theoretical background and impact mechanisms required for building a multidisciplinary evaluation methodology;
- Consider not only the expected positive impacts on road and driver safety and traffic flow, but also the unintended, and possibly negative, impacts of AD;
- Facilitate a good understanding of all possible effects of AD on the transport system, including the effects on equity of mobility and well-being of people, behavioural adaptation, safety and capacity, fuel consumption and emissions;

- Provide input to a Code of Practice for AD testing, interface design, and investigation of Human Machine Interaction (HMI).

To this end, the output of SP3 (WP3.3) will be a list of Research Questions (RQs), to be answered in the project, meeting the objectives defined above. This will be accompanied by development of innovative and appropriate experimental procedures to collect the data required to answer these questions, and the development of a structured and robust evaluation plan to ensure reliable and valid results are achieved from the pilot testing. To achieve this, L3Pilot will follow the **FESTA V-process methodology**, developed for planning, preparing, executing, analysing and reporting Field Operational Trials (FOTs) (Figure 1.2).

FESTA (Field opErational teSt supporT Action, 2007-2008) was a project that was set-up to produce comprehensive guidance on the evaluation and delivery of driver assistance systems and functions using a field operational test (FOT) methodology. The aim of the FESTA project was to provide a structured methodology that would ensure that the systems delivered real-world benefits. This matches a key objective of the L3Pilot project, hence the selection of this approach.

The FESTA Handbook (FOT-Net 2017) describes a process for evaluating driver assistance systems and functions. The four main pillars of this methodology will be followed in this project. These are:

(i) Prepare

(ii) Drive

(iii) Evaluate

(iv) Address legal and cyber-security aspects.

This process will be adapted to suit the needs of L3Pilot; taking into account the fact that the methodology has been developed for driver support systems, and pre-dates testing of most Level 3 automated driving functions. Therefore, the changes to this process will be documented as recommendations for a Code of Practise for the evaluation of AD functions (see D3.4 Final Evaluation Plan).

The SP3 methodology covers the steps in the left half (‘PREPARE’) of the updated FESTA ‘V’ (Figure 1.2), laying the foundations for the successful execution of the ‘DRIVE’ and ‘EVALUATE’ steps. The work in this sub-project (SP3) is done in close cooperation with other sub-projects. Recommendations given in the Trilateral Impact Assessment Framework for Automation in Road Transportation (Innamaa et al., 2017) are also followed. This work shows the complexity of the impact mechanisms resulting from indirect dependencies of key elements in the framework.

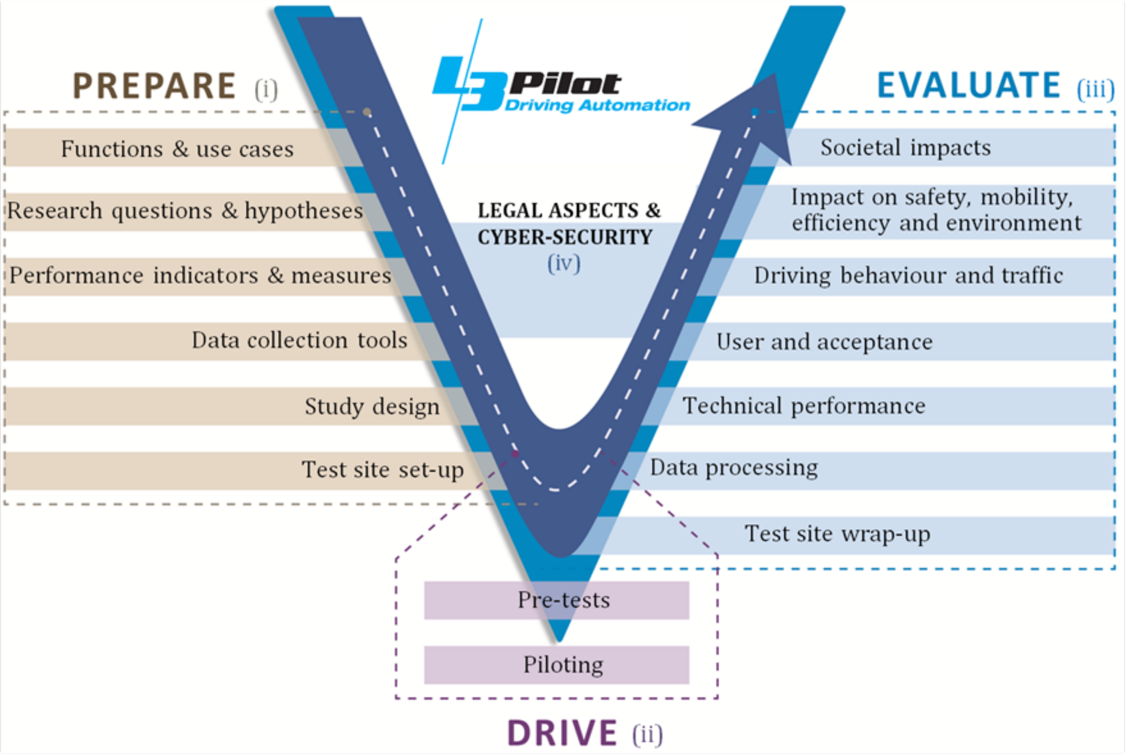


Figure 1.2: L3Pilot methodology overall structure

In L3Pilot, an evaluation of ADFs will be conducted to consider their technical performance, user and acceptance aspects resulting from knowledge of or interaction with the ADF, and driving and travel behaviour impacts of the ADF, see Figure 1.3.

The evaluation process will include data collection that involves logging subjective data from study participants and members of the public (e.g. questionnaires) and objective data from the pilot vehicles (e.g. vehicle CAN Bus). This data will be aggregated for the whole L3Pilot fleet. Based on the objective and subjective data from the pilots, the automated driving functions will be evaluated with respect to their technical- and user-related performance. Based on the findings (and the data) from technical- and user and acceptance evaluation, the impact of ADFs, in terms of their safety- and environmental effects, will be scaled up and evaluated, for all EU member states. Finally, the socio-economic impact of the three automated driving functions will be evaluated on a European level, in the form of a Cost-Benefit Analysis. This approach allows conclusions to be drawn of the likely impact of Level 3 (and Level 4) AD functions at the level of the individual driver, fleet, and also studying their effects, Europe-wide.

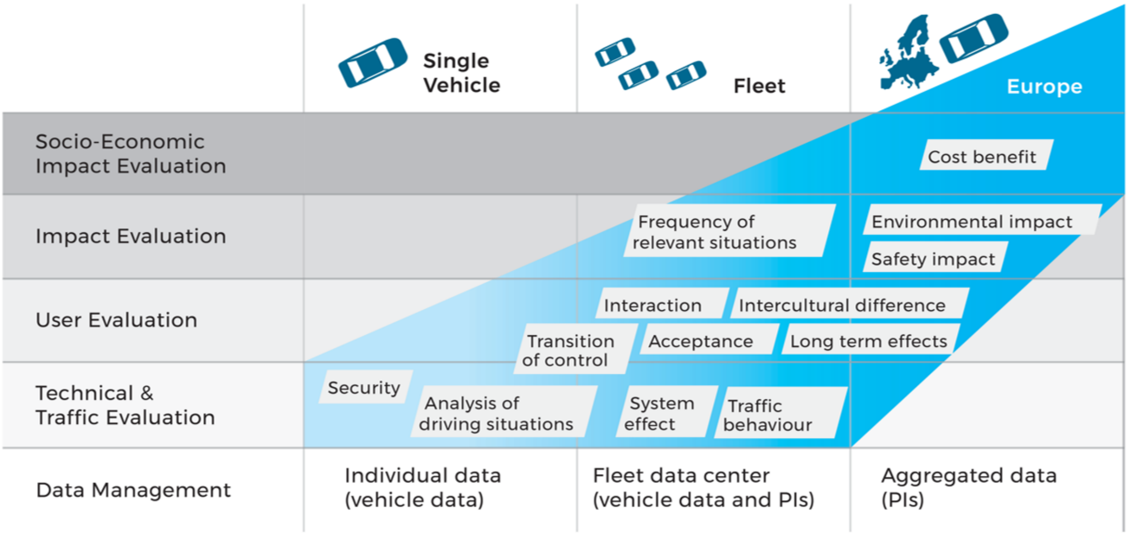


Figure 1.3: Process for evaluation in L3Pilot

1.3 WP3.3: Hypotheses and Indicators

WP3.3 focuses on the development of the Research Questions to be answered by the L3Pilot project. These Research Questions are then used to develop detailed hypotheses, before being linked to the performance indicators required for their assessment. Based on this list of performance indicators the required objective and subjective data logging needs in the vehicles and at the test site can be defined. This work supports the development of the experimental procedure (WP3.4), whereby experimental methods and pilot sites data are mapped on to the Research Questions, and evaluation methods (WP3.5), whereby suitable evaluation methods are sought for the Research Questions.

The approach of this work package is structured around the four key evaluation areas (Technical and Traffic, User and Acceptance, Impact, and Socio-Economic Impact) required for successful analysis and evaluation of the impact of the AD systems in SP7. The work package is sub-divided into two tasks (Research Questions and Hypotheses, and Performance Indicators, Measures and Data Logging), both of which are reported in this deliverable.

The definition of the ADFs from D4.1 (Description and taxonomy of AD functions) was used to inform WP3.3 work, such that RQs were developed with knowledge of the capabilities of the AD functions for testing.

This document presents the Research Questions formulated during the first seven months of the project (in March 2018). At this stage, an early assessment of RQ feasibility has been carried out using selection tools from previous projects and input from project partners on available methods, and data logging and simulation capabilities at each pilot site. For a final list, including details of the evaluation plan, reference should be made to D3.4: Evaluation Plan to be submitted in February 2020.

1.3.1 Task 3.3.1 Research Questions and Hypotheses

Task 3.3.1 defines and prioritises Research Questions (RQs) and Hypotheses for: (i) Technical and Traffic evaluation, (ii) User and Acceptance evaluation, (iii) Impact evaluation (safety, mobility, efficiency and environment), and (iv) for Socio-Economic Impact evaluation. RQs are generated hierarchically, first as high-level Research Questions, to provide a set of sub-categories for the development of more complex, detailed and specific questions. Hypotheses are then defined for the more detailed RQs, to support an extensive and detailed investigation of the impact of AD systems.

RQs for **technical evaluation** include aspects such as performance, reliability and readiness of ADFs, and security-related topics. Traffic and driving behaviour related RQs include topics such as car following, speed patterns, positioning in lane, and interactions with the infrastructure and other vehicles, e.g. behaviour at intersections.

RQs related to **user and acceptance evaluation** include aspects like trust, interaction with other vehicles and VRUS, transfer from the AV to the driver, human behaviour during AD, automation awareness, user acceptance (including willingness to buy and trust), use of ADFs, comfort and convenience. Possible age, gender, and cultural differences are also covered in this category of RQs.

The process of setting the RQs for **impact assessment** considers background literature and theory regarding the specific impacts that an ADF may have, detailed function descriptions considering the role of the user when operating them (developed in cooperation with WP4.3) and high-level use cases, taken from available ADF descriptions[2]. RQs for safety impacts cover direct and indirect impacts on road safety; for mobility impacts, the choice of travel mode and travel patterns are covered, for efficiency impacts, RQs cover road capacity and traffic flow, and incident frequency, and for environmental impacts, RQs cover emissions and energy use.

RQs related to **socio-economic** evaluation include different aspects of the economic and welfare effects of ADF use in Europe. Reduced car accidents, reduced mortality and health injuries are examples of socio-economic gains that can be measured in terms of saved statistical lives, quality-adjusted life-years, reduced hospital and rehabilitation costs, and less traffic disruption.

1.3.2 Task 3.3.2 Performance Indicators, measures and data logging

Task 3.3.2 specifies the **performance indicators** with which the Research Questions set in Task 3.3.1 can be answered, via the testing of their associated hypotheses. The performance indicators need to cover vehicle operation, interaction between the AV and other road users, driver behaviour, and subjective responses to ADF use. (Situational

---

[2] Use cases will be defined in more detail following precise specification of the ADFs in D4.1. This will be used to inform the development of specific RQs for each use case, which will be further reported in D3.4. Currently, the use cases are based on those outlined in draft versions of D4.1 and presented to the partners at SP4 meetings.

variables such as event type, traffic and weather conditions, and other map data may also be required to define a performance indicator). Task 3.3.2 ensures that all the specific data needs of the planned SP7 evaluation tools are met in the logging plan.

In addition, Task 3.3.2 specifies the necessary requirements for data logging and the steps needed to derive the performance indicators from this data. This also includes annotation of video data, which will include recordings of driver's eyes, head, body and foot movements, as well as their general activities in the vehicles, as well as recordings of the external environment.

2 Theoretical Basis for the L3Pilot Evaluation Framework

2.1 Overall evaluation framework

The L3Pilot evaluation framework follows the earlier mentioned FESTA V approach (Figure 1.2). This approach has been applied in the past mainly for the evaluation of technologies covering lower automation levels (mainly SAE Level 0 and 1) as well as for active safety functions and driver information systems. Examples include its use in the euroFOT (Malta et al., 2012), TeleFOT (Mononen et al., 2012), DriveC2X (Schulze et al., 2014) and interactIVe projects (Willemsen et al., 2011). L3Pilot evaluates automated driving functions which raises new challenges, although a first approach for the evaluation of automated driving functions (e.g. Level 2 systems) has been developed in the AdaptIVe Project (Rodarius et al., 2015).

An evaluation of automated driving is a complex task that needs to cover all different aspects of ADF use (e.g. safety, acceptance, environmental, socio-economic effects) and the related questions. It is essential that the L3Pilot evaluation framework follows a structured process to effectively deal with this complexity.

In order to structure the evaluation process and the approach to generating Research Questions, two dimensions have been defined in L3Pilot (see Table 2.1). The first dimension describes the **evaluation areas**. This dimension is process-oriented and indicates which work packages of the evaluation sub-project (SP7) deal with the evaluation of an aspect of ADF use. This structure is in line with previous research projects – starting from PReVAL (Scholliers et al., 2007) up to AdaptIVe (Rodarius et al., 2015). In particular, this dimension covers four different evaluation areas:

i) Technical and Traffic Evaluation,

ii) User and Acceptance Evaluation,

iii) Impact Evaluation, and;

iv) Socio-Economic Impact Evaluation.

The second dimension of the evaluation framework is the **impact areas**. These impact areas include system performance, driver behaviour, user experience, mobility, safety, efficiency, environment and socio-economics (Table 2.1). They are in line with previous projects like TeleFOT (Mononen et al., 2012) and DRIVE C2X (Schulze et al., 2014) as well as the recommendations of the Trilateral Impact Assessment Framework (Innamaa et al., 2017). The impact areas provides an additional dimension of categorisation for the developed research questions, and so helps structure the evaluation stage of the project.

Table 2.1: Mapping of Evaluation areas and Impact areas; x = Evaluation area under which the evaluation is made, (x) = Evaluation area affected by the results

		Evaluation Area			
		1. Technical & Traffic Evaluation	2. User & Acceptance Evaluation	3. Impact Evaluation	4. Socio-Economic Impact Evaluation
	• System Performance	X		(X)	
Impact Area	• Driver Behaviour	X		(X)	
	• User Experience		X	(X)	(X)
	• Mobility		X	(X)	(X)
	• Safety	(X)		X	(X)
	• Efficiency	(X)		X	(X)
	• Environment	(X)		X	(X)
	• Socio-economics				X

The adopted four evaluation areas are described in more detail in the next section.

2.2 Technical and traffic evaluation area

This evaluation area includes an analysis of System Performance and Driving Behaviour.

2.2.1 System Performance

System performance involves analysing the technical performance of the L3Pilot AD functions[3]. The aim of this analysis is to understand the technological readiness of the systems, and to provide the necessary input data for further analysis in the Impact Evaluation and Socio-Economic Impact Evaluation process. The main focus of the analysis of system performance is to understand how reliable the function is in a given driving and traffic scenario, and under what circumstances it needs to issue a takeover request to the driver. To be able to model the behaviour of the function in a simulation (e.g. driven velocity, accelerations, distance while driving in automated mode), data also needs to collected about the potential exposure of the AD function (when, where and in which conditions it is available) as well as a description of the general driving behaviour when using the system.

---

[3] Note that the ADFs are currently under-development but are expected to be at a sufficient level of maturity by the experimental phase, to allow meaningful scaling up of the results of the pilot studies.

Furthermore, potential risky situations, such as violations of traffic rules or any critical driving situations that were caused by the function should be identified.

One important aspect is that the analysis does not focus on single OEM-specific performances, but provides an overview of the general performance of automated driving functions according to their automation level and operational design domain (ODD). This is a necessary approach to take to protect vehicle owners, however, where possible specific analyses will be conducted in an anonymised fashion, such that individual ADFs cannot be identified from the findings.

2.2.2 Driving Behaviour

In the context of automated driving, the concept of 'Driving Behaviour' expands from the 'typical' driving behaviour of a human, because, here, driving behaviour includes both vehicle behaviour (when the AD function is switched on) and (human) driver behaviour, as well as the likely hybrid scenario, where they may be momentary transitions of control between the two.

To observe any impact of the AD function on safety, mobility, efficiency, and environment, some noticeable change must be realised in the behaviour of the driver or vehicle. This statement is specifically relevant to pilot studies like L3Pilot, which are designed to collect new data, and produce findings regarding the first order measures (driving behaviour; as part of the Technical and Traffic Evaluation) before a conclusion is drawn about the second order impacts (such as safety; as part of the Impact Evaluation). Driving behaviour is assessed through the pilot studies, whilst impact evaluation involves using this pilot data in modelling processes or expert assessment. These noticeable changes are also important when trying to understand impact in terms of monetary values for socio-economic assessment of ADs. The schematic picture in Figure 2.1 presents the hierarchy of different types of impacts.

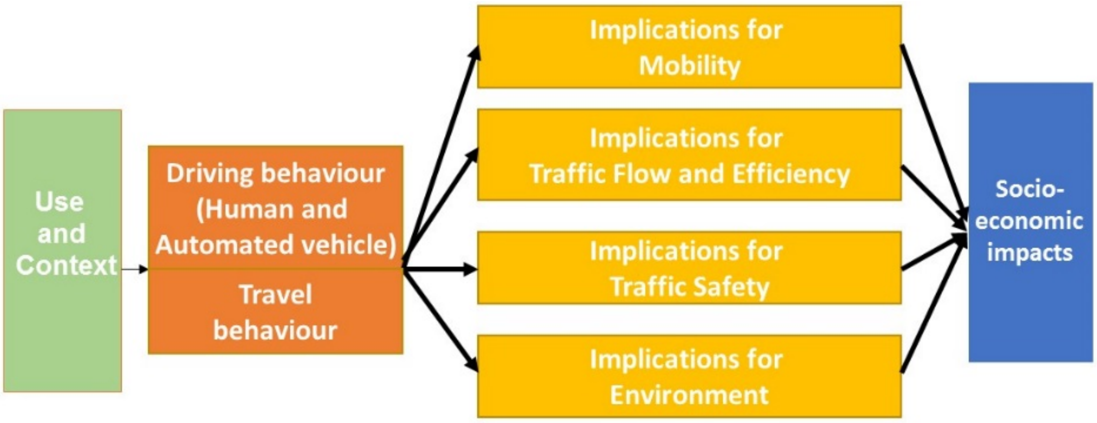


Figure 2.1: Schematic picture of the hierarchy of different types of impacts

The Trilateral Impact Assessment Framework (Innamaa et al., 2017) defines driving behaviour (in this context, vehicle-controlled or driver-controlled driving behaviour) as:

acceleration, deceleration, lane keeping, car following, lane changing and gap acceptance. Relevant automation applications include those which provide longitudinal and/or lateral control with respect to the road and other vehicles.

2.3 User (interaction) and function acceptance evaluation area

When assessing impacts of an ADF on a transport system, changes in user behaviour that may be observed as they interact with the function are essential. Moreover, the users need to accept the function and use it, to allow any impacts of this use to be evaluated. Therefore, a good understanding of the user, their interactions with the ADF and their opinions about it has a central role in the evaluation process. Examples of such user observations may include, when and why the AD function is engaged or disengaged, whether and when drivers trust the system (e.g. by using subjective measures, or observing its deliberate use) and how an AD function might change drivers' response to unexpected events in the environment.

2.3.1 User experience

The driving task of controlling the vehicle can be described at three levels, which include the strategic, the tactical, and operational control of the vehicle (Michon, 1985, Molen and Botticher, 1987). Some examples of activities related to each level are presented below:

- **Strategic level**: trip planning, minimising time, avoiding traffic and route finding
- **Tactical level**: interaction with other road users, selection of driving speed and distance to the others (including, for example, gap acceptance), deciding to change lane.
- **Operational level**: vehicle control on the road, steering and braking.

This classification provides a systematic structure for studying different aspects of driving behaviour and was kept in mind when formulating the Research Questions. In the methodology developed for L3Pilot, the **strategic level** equates to 'travel behaviour' and relates to '(personal) mobility'. The strategic decisions are typically the most critical for the wider impacts on safety, mobility, environment and efficiency, as they influence the whole journey. The **tactical level** is directly linked with driving behaviour through speed and headway choices. With the increased degree of automation, the concepts which have originated from the work domain are now emphasized in road transport. Examples of these concepts are Situation Awareness (Endsley, 1995) and workload, which have been studied in terms of the effect of automation on traffic safety (e.g. De Winter et al., 2014, Stanton & Young, 2010). Automation awareness in another concept considered here, and refers to how well the user is aware of the capabilities and constraints of an automated system, either in terms of a general knowledge about the function, or in the actual situation (Whitlow et al., 2002). The **operational level** is linked to the technical performance of the ADF (or the driver's performance when the ADF is not switched on), and refers to moment-by-moment control of the vehicle.

2.3.2 User acceptance in the long-term

In order to predict user behaviour, it is important to understand if and how their interaction with the automated system will change with experience, and how this behaviour is affected by system limitations. Behavioural changes after long-term use of ADAS are illustrated in a model by Martens and Jenssen (2012). The learning phases in the **Behavioural Adaptation** to ADAS are defined as: *First Encounter, Learning, Trust, Adjustment and Readjustment*, and are expected to last up to two years. These changes can occur in terms of changes in drivers' perception, cognition, performance, state and attitudes towards the ADAS.

User acceptance or behavioural adaptation to driving automation is a process that develops over time and with exposure. Therefore, repeated long-term trials are necessary to get a more comprehensive understanding of the interaction of drivers with automation and their attitudes towards automation. There can be an immediate effect on user acceptance and user behaviour at the time of the first encounter, for example, within the first hour of use. This is also the time span that is usually considered in user studies. However, it can be expected that with longer usage the user behaviour towards the system may change again due to experience of more driving situations (Manser, Creaser & Boyle, 2012).

2.4 Impact evaluation

The impact evaluation involves consideration of the wider impacts of the ADFs in a number of impact areas. For this evaluation, the focus is not on individual system performance, driving behaviour, and user acceptance, but instead on the higher level assessment of ADF impact overall, using data combined across many the pilot sites. The methods used to combine data for these analyses are the subject of subsequent tasks in SP3 (see D3.3 Evaluation Methods).

2.4.1 Safety

The three relevant aspects to be covered in traffic safety impact assessment are (1) exposure, (2) crash risk, and (3) consequences in a crash. Traffic safety is regarded as a multiplication of these three orthogonal factors (Nilsson 2004, Figure 2.2).

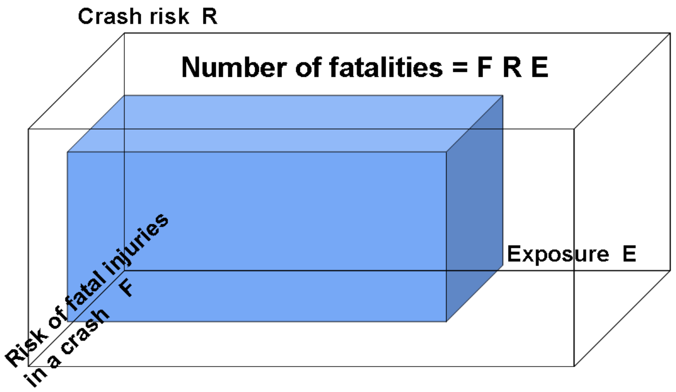


Figure 2.2: The dimensions of road safety (Nilsson 2004)

Based on the work conducted by Draskóczy et al. (1998), Kulmala (2010) and Kulmala et al. (2008) suggest nine impact mechanisms that need to be considered when studying the traffic safety effects of ITS. These include:

1. Direct in-car modification of the driving task
2. Direct influence by roadside systems
3. Indirect modification of user behaviour
4. Indirect modification of non-user behaviour
5. Modification of interaction between users and non-user
6. Modification of road user exposure
7. Modification of modal choice
8. Modification of route choice
9. Modification of accident consequences.

Mechanisms 1 to 5 deal with crash risk (Kulmala, 2010). The Research Questions related to crash risk relate to: 'Speed', 'Proximity', 'Position', 'Interaction', 'Use of signals', 'Driver condition', and 'Attention'. The following assumptions concern mechanisms 1–5 Kulmala, 2010; Nilsson, 2004, Gelau et al., 2008), with safety assumed to increase as:

- speed decreases (power model), Nilsson (2004)),
- standard deviation of speed decreases,
- number of jerks decreases,
- speed violations decrease,
- very close following decreases,
- lateral position is more stable,

- vulnerable road users are given consideration,
- signals are used correctly,
- driver state is not deteriorated,
- the focus of driver attention is allocated correctly.

Mechanisms 6, 7 and 8 are related to exposure. Accordingly, the related RQs regarding exposure include variables such as (1) 'Time spent on road' (2) 'Mode chosen for the journey', (3) 'Timing of the journey', and (4) 'Road type used'. 'Time spent on road' has a linear relationship with safety, where traffic safety decreases as mileage increases (see Elvik et al., 2009). Choice of transport mode has relevance as well, since, for a given mileage, public transport is safer than driving private cars (Elvik et al., 2009). Timing of journey affects traffic safety, because driving during peak hours and night-time is more dangerous than driving at other times (Elvik et al., 2009). There are also differences in crash risk between different road types, showing lower risk for motorway driving (Elvik et al., 2009). Consequently, safety increases as the proportion of motorway driving increases, and safety increases as the proportion of urban driving decreases.

Finally, two assumptions are related to the modification of accident consequences Mechanism 9). First, it is assumed that the consequences are likely to be more severe as speed increases (fatality equation, Kulmala, 2010), and secondly, there is a relationship with vehicle type, as bigger, newer and more robust vehicles provide a higher level of safety and are more protective in a crash. However, it should be noted, that this refers to modification of accident consequences for the driver of the vehicle. Enhanced vehicle safety systems may lead to reduced accident consequences for other road users e.g. pedestrians.

2.4.2 Mobility

The term 'mobility' is often used to describe travel, without further definition, and is used in a number of different contexts (Carlson & Marchi, 2014; Metz, 2000). In transport research, personal mobility refers to movement of people through geographical space and time. However, mobility is more than just travel or transport. It is defined as the ability to move (Hanson, 1995), the ease of movement (Sager, 2006), or the potential for movement (Spinney et al., 2009; Gudmundsson, 2005). Furthermore, some definitions emphasise that mobility includes people's preferences of travel, their feelings, and their decisions over time, route and mode (Hakonen 2010; Button et al., 2006; Gudmundsson, 2005). Thus, "*mobility is not just a matter of where one can travel to but also entails the ease of travel. In many cases, it is the quality of travel that is more important than the simple ability to get somewhere*." (Button et al., 2006, p.19). Mobility is also based on mobility tools, such as the networks and means of travel one knows about, has access to, and is willing to use (Kulmala & Rämä, 2010; Spinney et al., 2009). Therefore, revealed mobility (in other words, the benefits derived from travel activities) happens within mobility (Spinney et al., 2009).

Based on these references, mobility in L3Pilot is defined as the potential for movement of people. It consists of means of travel and networks one has access to, knows about and is willing to use. Along with transport and infrastructure it encompasses people's and road users' intentions, opinions and choices in their daily travel and movement. Since measuring mobility is difficult, revealed mobility can be used as an imperfect measure of it (Innamaa et al., 2013; Spinney et al., 2009). Revealed mobility refers to the movement of people (travel). In L3Pilot revealed mobility as a measure can be complemented with other measures aiming to capture movement potential more comprehensively.

A mobility model, developed for the TeleFOT project (Innamaa et al., 2013), and used in the DRIVE C2X project (see also Malone et al., 2014), provides a structure for revealed mobility. In the model, revealed mobility consists of the amount of travel (number of trips, length and duration), travel patterns (timing, mode, route and adverse conditions) and trip quality (user stress, user uncertainty, feeling of safety and feeling of comfort). Of these, amount of travel and travel patterns can be measured by objective means, whereas trip quality is a matter of subjective experience.

Even if mobility is hard to capture completely, it is an important measure when assessing the impacts of future transport phenomena, such as automation of cars. These future impacts are hard to assess by using descriptive data of past trips alone. Mobility (and so also revealed mobility) consists of multiple environment-related and personal factors (Kuisma, 2017). Environment-related factors are, for instance, availability of travel modes and services, the design of transport environments, traffic situations and weather conditions (Klöckner & Friedrichsmeier, 2011; van Wee et al., 2002). Personal mobility-shaping factors include demographic and socioeconomic background, life situation, personality, identity and preferences (Kuisma, 2017; Paulssen et al., 2014; Murtagh et al., 2012). Mobility relates to the strategic level of driving and travelling tasks since needs, resources and habits of an individual shape strategic level activities.

Kuisma (2017) grouped the resources to travel into four categories: time, money, physical abilities and mental resources. Time means the time an individual has available to use for travelling. Money refers to the monetary resources one has. Physical abilities refer to the physical capability, skills, competence and energy required from an individual to travel. Mental resources, for their part, refer to the mental and cognitive capability, skills and competence, as well as mental energy or coping required from one to travel. The limits of these resources restricts and shapes mobility and thus travel: it can impact both the amount and patterns of travel as well as the quality of it. Different resource constraints are crucial for different people.

2.4.3 Efficiency

Traffic efficiency describes how efficiently (in terms of average speed and travel time, number of stops, delay etc.) people and goods can move through the transport network. The number of vehicles passing through a cross-section of a road, during a certain time, is called traffic flow (also traffic volume or throughput). The capacity of a road is defined by the maximum traffic flow, that is, the maximum number of vehicles that can pass by a point on the road in a period (e.g., 1 hour).

Traffic flow links traffic volume, density and speed. The traffic volume is equal to traffic density times speed. On the other hand, traffic density is influenced by the headway selection.

Capacity is influenced by several factors and their interactions (see Figure 2.3), including: environment (such as the layout of the road or weather conditions), vehicle characteristics (such as vehicle type, ADAS and/or ADFs), driving behaviour (such as preferred headway and driver state) and mobility behaviour (such as selection of travel mode, and thus resultant impacts on vehicle usage rates). Impacts on traffic flow depend on the penetration rate of automated vehicles, and changes in travel behaviour, but also on the potential regulations regarding, e.g. car following behaviour.

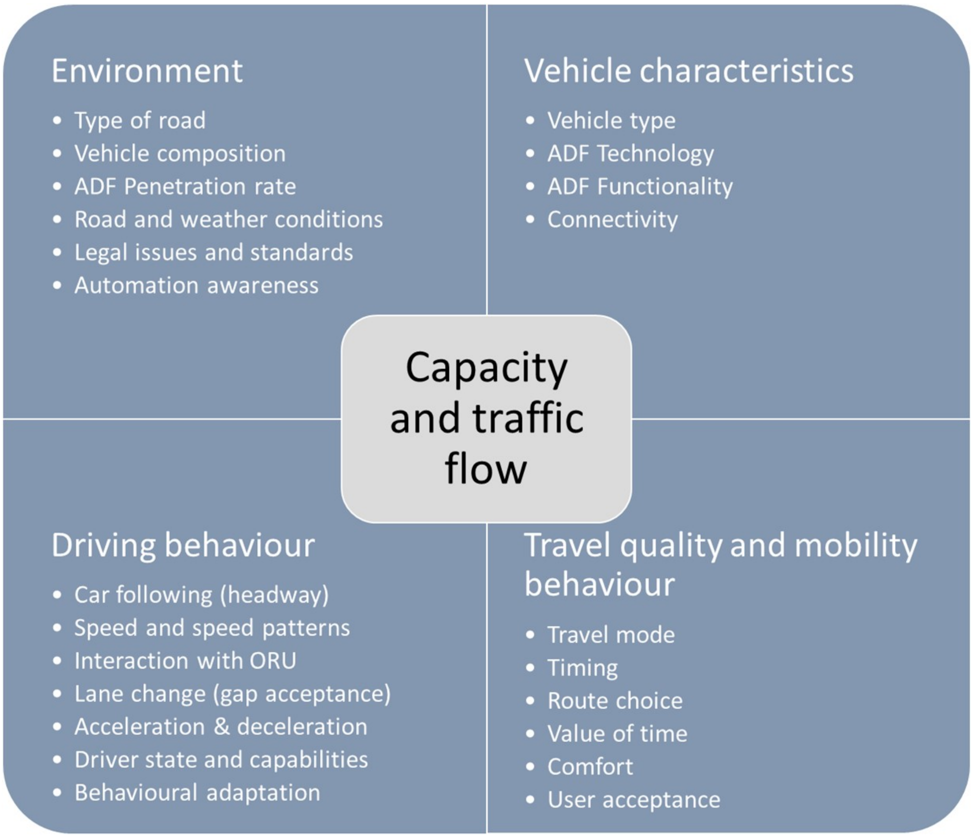


Figure 2.3: Transport network efficiency impact factors

Traffic volume or the traffic demand is affected by changes in mobility habits: number and timing of journeys, mode of travel, vehicle occupancy, route and road choice. These decisions are made at the strategic level. On a higher level, these choices are affected by land use, transport system design, as a whole, and demographics.

Changes at the tactical driving level, such as changes in speed, acceleration and deceleration as well as headway to other vehicles, and gap acceptance affect efficiency of traffic flow. Thus, traffic flow is affected by vehicle operational factors. These are tasks that have typically been carried out by drivers, but that are being more and more taken over by automated driving functions (Innamaa et al., 2017).

Research conducted so far provides mixed conclusions on the effects of automated driving on roadway capacity. Conclusions vary from slight reductions in capacity to large increases (Department for Transport, 2016). The results largely depend on the initial assumptions made, and are based on models and estimations only. Traffic efficiency is affected by many different, but interrelated factors and the development of those factors are still uncertain. According to Milakis et al., (2017), the benefits of automated driving on traffic flow efficiency are highly dependent on these factors, which include:

- Level of automation
- Connectivity between vehicles
- Penetration rates
- Deployment path (dedicated lanes or integrated in mixed traffic)
- Human factors (behavioural adaptation; appropriate trust and use)
- Changes in demand (increased demand possible).

However, the long-term implications of automation are currently uncertain and largely depend on the development of travel demand, as well as the business models adopted: for example, whether the vehicles will be personal or shared.

Automated driving functions are expected to provide a new and interesting interplay between safety, comfort and efficiency. For example, due to shorter reaction times to impending collisions, automated vehicles are potentially able to drive with shorter headways. On the other hand, due to safety or comfort reasons, the headways chosen may also be longer than those currently chosen by human drivers. Therefore, it is likely that some trade-offs will be made between those impact areas (Milakis et al., 2017).

Studies suggest that in the initial phase, with a low penetration rate of AVs, road capacity may decrease due to AVs behaving cautiously in the presence of human-driven vehicles (Milakis et al., 2017, Department for Transport, 2016). This implies that there is a conflict between capacity and safety, and capacity and traveller comfort. These conflicts may take on a different form than those that exist currently for manual driving. For the sake of user acceptance of AVs, (and to maximise safety and comfort, and minimise liability issues), the vehicles may first be configured to use larger headways and lower acceleration rates than those adopted by average human drivers. This may have adverse effects on the roadway capacity. Thus, a vehicle that is designed for the comfort of its users (allowing to use the travel time for other activities) can have unintended impacts on roadway capacity for other road users (Department for Transport, 2016).

Currently, it is not yet clear how the parameters used by the ADF will be defined in the long term: by OEM, user preferences, legislation, or some combination of these. In addition, more research is required into how users may be willing to trade an increase in travel time against the ability to use the available time in a more efficient and convenient way (Department for Transport, 2016).

The interaction between AVs and manually driven vehicles is also an important aspect to study with respect to traffic efficiency. As automated vehicles are being introduced gradually into the current transport system, there will be a period of time with mixed traffic on the roads. AVs are expected to confront difficulties in the conventional traffic system in several ways (Calvert et al., 2016):

- Anticipatory capability of AVs is not as good as of a humans, AVs act more reactively rather than proactively
- Behavioural recognition of AVs is limited (i.e. how they react to different traffic situations)
- AVs have limited flexibility
- AVs miss human courtesy and are non-sociable
- AVs may be treated differently to conventional vehicles, in terms of their equality or status in traffic

On the other hand, human drivers also face challenges with increasing automation of vehicles. There is a range of factors contributing to the cause of automation-related interaction errors and accidents, including insufficient or inappropriate system feedback, misunderstanding of automation capabilities, and over-reliance on automation (Billings, 1997; Parasuraman and Byrne, 2003; Parasuraman and Riley, 1997). In addition, a number of inter-related psychological factors have been linked to drivers' capacity or willingness to interact safely with automated driving systems, including trust (Lee and See, 2004; Hergeth et al., 2016), locus of control (Stanton and Young, 2005), complacency (Bagheri and Jamieson, 2004; Parasuraman and Manzey, 2010), mental models (Moray, 1990; Sarter et al., 2007; Flemisch et al., 2012), driver state (Rauch et al., 2009; Neubauer et al., 2012; Jamson et al., 2013; Louw & Merat, 2017), mental workload (MWL; De Waard, 1996; Collet et al., 2003), and situation awareness (SA; Endsley, 1995a; Merat and Jamson, 2009; Kircher et al., 2014).

Several studies suggest that connectivity and cooperation between vehicles, and vehicles and infrastructure, is essential for any benefits to traffic flow. It is expected that traffic throughput will suffer if AVs are introduced before sufficient implementation of connectivity is present (Calvert et al., 2016). Even with connectivity, a significant penetration rate of connected AVs (about 40 %) is needed to achieve significant impacts (Milakis et al., 2017, Department for Transport, 2016).

2.4.4 Environment

The impact of automated driving on the environment depends considerably on travel, driving conditions, and driver and vehicle behaviour. The environmental impacts of traffic include, for example, greenhouse gas emissions, particle emissions and noise. The factors, which are expected to play a role in the impacts of ADFs on emissions, noise and surroundings are shown in Figure 2.4. Advances of technology, such as different engines, should also be taken into account.

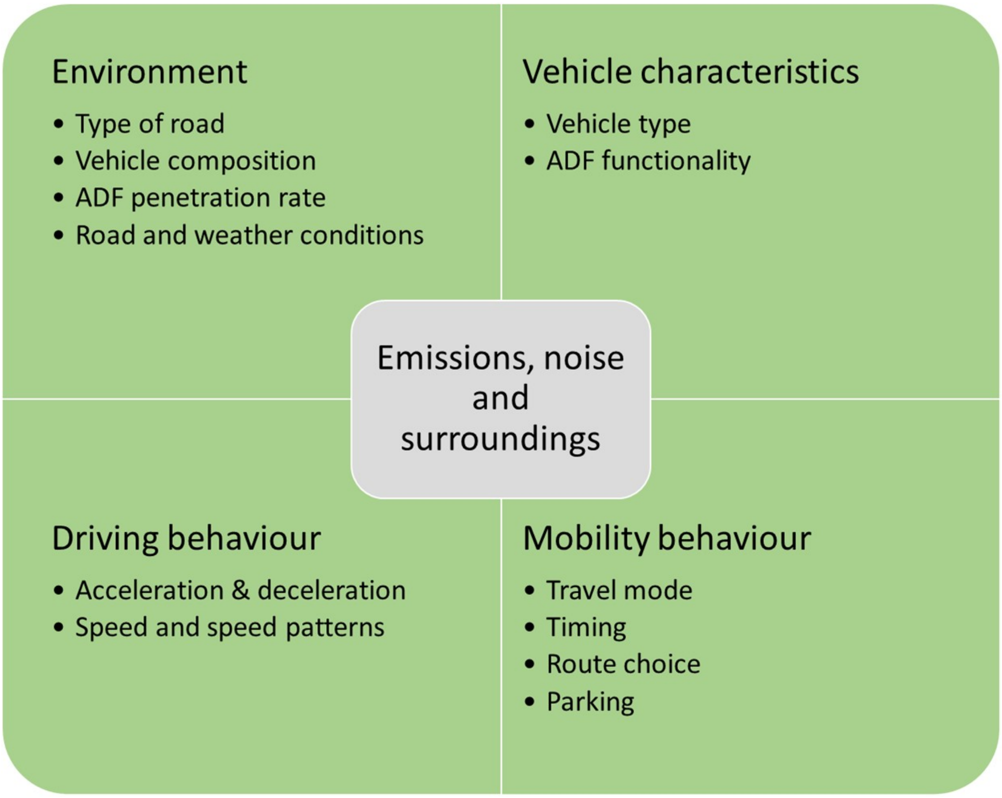


Figure 2.4: Environmental impact factors

Noise is a multifactorial effect and depends, for example, on the type of road (pavement), type of vehicle, tyres, and driving manoeuvres (e.g., steady state, acceleration). The following assumptions concern the assessment of the noise impact of AVs:

- Number of journeys: There is a correlation between number of journeys and noise, where, essentially, each avoided journey means a reduction of noise.
- Route: Changing the route may reduce the negative effect of noise if there is a shift of traffic from rural and urban roads to motorways. This is not an absolute change, but a qualitative one, because normally motorways are located in areas away from residential areas, and in many places there are noise barriers, when the distance is short.

- Speed: There is a direct relationship between speed and noise. Noise increases when vehicle speed increases. However, there is wide variability across vehicles, and no precise equations can be used, although the FHWA has developed a Traffic Noise Model (see FHWA, 2011) which considers speed as a parameter.

Despite the great results achieved in the recent years regarding emission reduction, vehicles still contribute to emission-based pollution (OECD, 2017). There are several factors, which influence emissions, depending on travel and driver behaviour:

- Number of journeys: There is a direct link between number of journeys and emissions.
- Travel mode: There is a positive effect (reduced emissions) by using public transport, bicycle, walking or multimodal transport, instead of using private vehicles only.
- Length/duration: There is a proportional link between travel duration/length and emissions.
- Time budget/timing: Changing travel time may mean different traffic conditions encountered, which has an impact on emissions.
- Route: Changing route may influence emissions. For example, the opportunity to use an urban shorter road without traffic jams, supported by an intelligent traffic light system, may be a positive alternative to a longer high-speed road.
- Speed: The relationship between speed and emissions is different for the different pollutants. Nevertheless, considering a vehicle running at constant speed, if this speed is low the vehicle has high emissions. For speeds somewhat higher, the emissions decrease in a smooth way as a function of this speed level. For highest speed, the emission vs. speed curve rises again due to aerodynamic effects. However, in real traffic conditions there are variations in speed, which also affect emissions. For this reason, in order to evaluate the impact of ADFs, it is important to consider not only the mean speed, but also the standard deviation of speed. Alternatively a comparison of the complete speed profile may also be useful.
- Pedals: The way the accelerator and brake pedals are operated is important for emissions. Rapid changes normally imply high emissions.
- Fuel consumption: There is a linear relationship between fuel consumption and CO2 emissions. However, the relationship is more complex for other pollutants, which depends on the particular transient condition encountered.
- Surroundings: Changes in exposure bring changes in emission and noise levels, and therefore the attractiveness of surroundings

2.4.5 Direct and indirect effects

The potential effects of automated driving have been widely discussed, and are linked to both direct and indirect effects (see e.g. Milakis et al., 2017 or Innamaa et al., 2017). As defined by Innamaa et al. (2017), **direct effects** are often related to safety, vehicle operations, environment or personal mobility. These can be measured, for example, via traffic conflicts (safety), driving behaviour, car following and intersection performance (vehicle operations), energy consumption and emissions (environment), and the comfort of the user or the users' ability to multi-task while in the vehicle (personal mobility). Direct effects of automated driving are typically (though not always) short-term in nature and therefore these effects can be measured in field operational tests.

Indirect effects are commonly implications of the direct effects (i.e., changes in travel and road user behaviour). Indirect effects of automated driving are, therefore, typically more difficult to measure than the direct effects. For one reason because their time horizon is longer. This is also the reason why indirect effects cannot be easily measured in field operational tests or in a pilot study. The indirect effects develop over time, and are produced as a result of a path/chain of impacts, often with complex interactions and external factors (Innamaa et al., 2017). For example, from a safety point of view, the adoption of riskier behaviour because of over-reliance on the system, may have indirect effects on traffic safety. Therefore, as indicated by Milakis et al. (2017), traffic safety can improve in the short term, after the introduction of automated vehicles, but behavioural adaptation (and low penetration rates of vehicle automation) might reduce these benefits.

The separation of potential effects of automated driving into direct and indirect effects is also supported by Milakis et al. (2017), who propose a model to conceptualise the sequential effects that automated driving can bring to several aspects of mobility and society. The model consists of impacts of automated vehicles at three different stages: first (i.e., direct), second and third order (indirect) effects. The model illustrates the implications automated driving is expected to have, first on traffic, travel costs and travel choices (first order direct effects), and then affecting second and third order indirect effects such as mobility, environment, efficiency, safety, and public health.

According to Milakis et al. (2017) most studies show that automated vehicles could induce an increase of travel demand between 3% and 27%, due to changes in destination choice (i.e., longer trips), mode choice (i.e., modal shift from public transport and walking to car), and mobility (i.e., more trips). In addition to indirect effects to traffic safety, these changes can according to Milakis et al. (2017), have indirect effects also on efficiency and environment.

It is important to note that, as reported by Milakis et al. (2017), the sequential order of the effects of automated driving is not always straightforward. Some feedback can occur from higher-order (indirect) effects to lower-order (direct) effects.

Kulmala's framework for the safety assessment of ITS (2010), is based on a nine-point list of ITS safety mechanisms, which emphasises the direct and indirect modification of driver behaviour (see Section 2.4). Furthermore, the mechanisms focus on changes in the mobility of road users, since changes in exposure, modal choice, and route choice can potentially have indirect effects on traffic safety. Innamaa et al. (2017) suggest a similar approach for systematically assessing the indirect impact of user and non-user behaviour, as well as taking into account factors such as: changes in the interaction between automated vehicle and other road-users, exposure / amount of travel, modal choice, route choice and consequences due to different vehicle design.

The impacts paths of automated driving for different impact areas such as safety, network efficiency and personal mobility are depicted in Figure 1.8. This figure describes through an example the potential paths starting from direct impacts on vehicle operations, driver or traveller, to the quality of travel and transport system. Naturally, strong links can be found between the impact areas (e.g. safety impacts affecting efficiency and environment) which support the importance of the assessment of indirect effects.

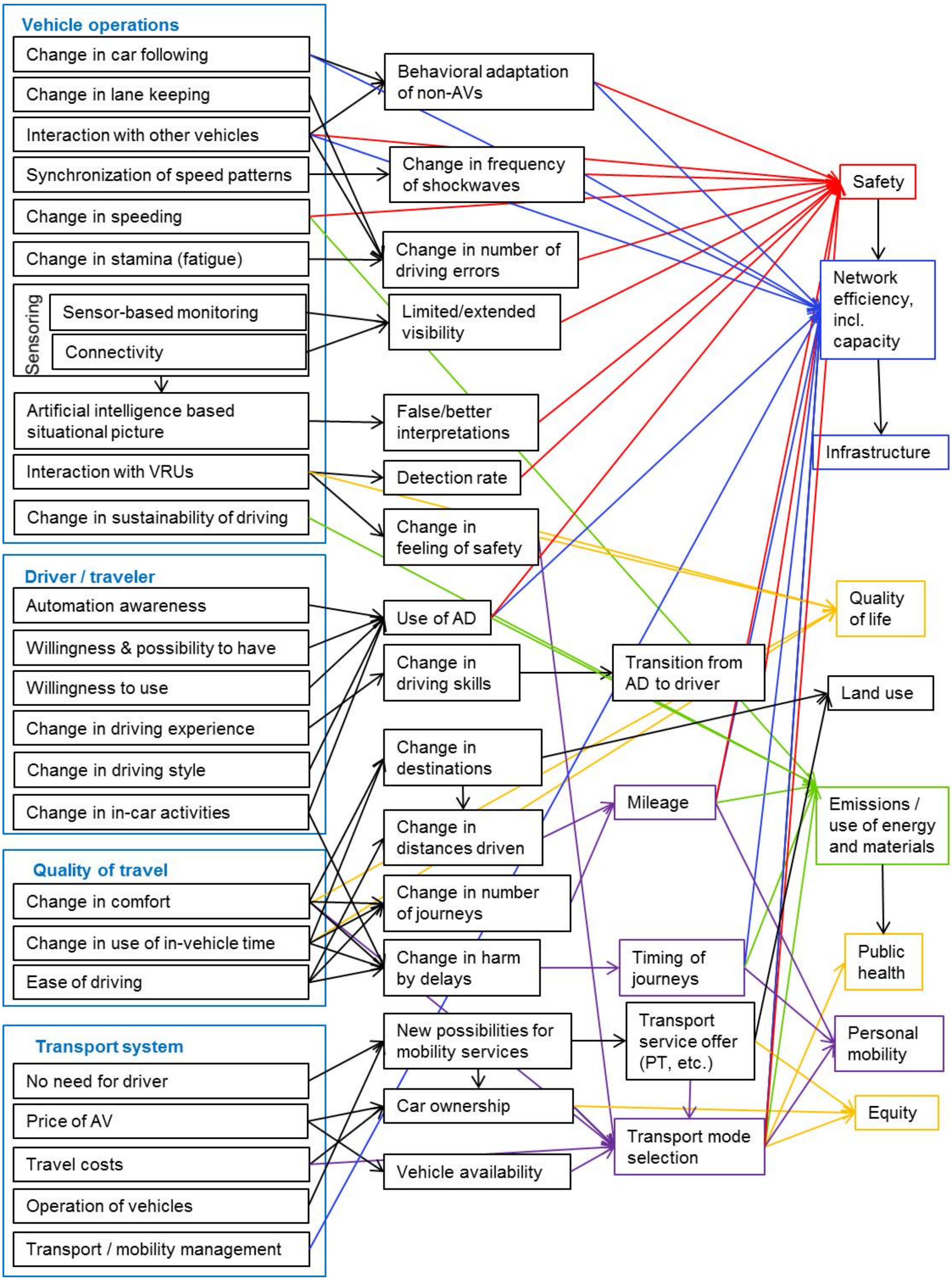


Figure 2.5: Impacts paths of automated driving for different impact areas (Innamaa et al., 2017)

2.5 Socio-economic impacts

2.5.1 Cost-benefit analysis

In transport projects, economic appraisal is carried out to measure the magnitude of the economic impact resulting from the investment. In an ideal case, the total benefits arising across all final product sectors would be measured, including changes in the labour market, prices of goods and services in the market, as well as value of property in the land market. However, such an analysis would require a level of sophistication, which is typically beyond that available from technical and resource standpoints (HEATCO, 2006).

An established methodology for comparing costs of measures with the resource savings, such as increased traffic safety, reduced travel times, and fuel consumption, is socio-economic cost-benefit analysis (CBA). Socio-economic cost-benefit analysis (CBA) provides a quantitative measure of the extent to which certain measures provide societal benefits that exceed the costs of their implementation and operation (ITF 2017). This assessment can be undertaken by individual stakeholders, by using the costs and benefits that are relevant to them, for example in assisting with the development of business cases. More commonly, a CBA is undertaken at a societal level (Stevens et al., 2016).

At the societal level, CBA is based on welfare economics, where resource savings make up benefits because of the assumption that those resources could be used in other parts of the society with at least the same productivity. CBA is used as a measure to compare the costs of a certain investment with the resource savings, or benefits expected over a certain time period. Typically, the impacts considered relate to safety, mobility and the environment, as these are the most readily monetised with standardised cost-unit rates. In addition, the costs of the services have to be estimated in terms of investment costs, operation costs and maintenance costs. Costs and benefits are calculated over a defined period. CBA is often complemented by sensitivity analysis, by varying parameters of the CBA calculation (Stevens et al., 2016). Socio-economic CBA is typically carried out to compare two alternative future deployment scenarios: a *base case* (“do-nothing-scenario”), which assumes that no services are implemented, and a *with-case* (“do-something-scenario”), where new services are implemented (Stevens et al., 2016).

A comprehensive guide to CBA analysis of ITS in general is provided by the European Commission’s “Guide to Cost-Benefit Analysis of Investment Projects” (2014). Stevens (2004) summarizes the typical phases of typical ITS cost-benefit analysis as follows:

1. Define assessment objectives; intended effects of ITS, such as efficiency, safety, environmental effects; timescale of the evaluation
2. Describe system characteristics; technology included, performance, expected market penetration

3. Define assumptions concerning policy and technology; mandatory/optional use and public acceptability/use, legislative and standard issues, maturity of technology and alternatives
4. Identify impacts; expected impacts to be included into the assessment
5. Select appropriate indicators; e.g. time savings, injury accidents, vehicle operating costs, passenger comfort, etc.
6. Estimate effects on indicators, e.g. vehicle hours driven/saved, number of accidents, etc.
7. Apply monetary unit values; e.g. national unit values for e.g. injury/fatal accident, vehicle hours driven/lost due to congestion, CO2 emissions etc. or if not available, consider using common European values.
8. Perform analysis and present results.

2.5.2 Wider societal impacts

Investments in transport improvements are often reasoned by their impact on economic performance: it is assumed that investments act as a catalyst for private sector investment. Transport investments can also deliver economic benefits beyond conventionally measured user benefits. These wider economic impacts typically reach further than conventional transport CBA, which focuses on the user benefits created by a project and often assumes no change in land use (ITF, 2017). The wider impacts are relevant especially in the context of automated driving, which is expected to lead to wide-reaching societal effects (Milakis et al., 2017).

As stated above, time and cost savings are the core of transport appraisal. They are illustrated in the left part of Figure 2.6 These changes are often referred to as the user benefits of a project. The International Transport Forum (2017) expands this view by including wider economic impacts, seen in the right hand part of the figure. They arise as a consequence of transport impacts on economic geography: improved transport options increase perceived proximity of economic agents and may also trigger relocation of economic activities. The potential for wider economic benefit is created by three main mechanisms (ITF, 2017):

- Proximity and relocation shape effective density of economic activity, and thus productivity
- Transport improvements make affected locations more attractive for investments
- Impacts on the labour market may occur

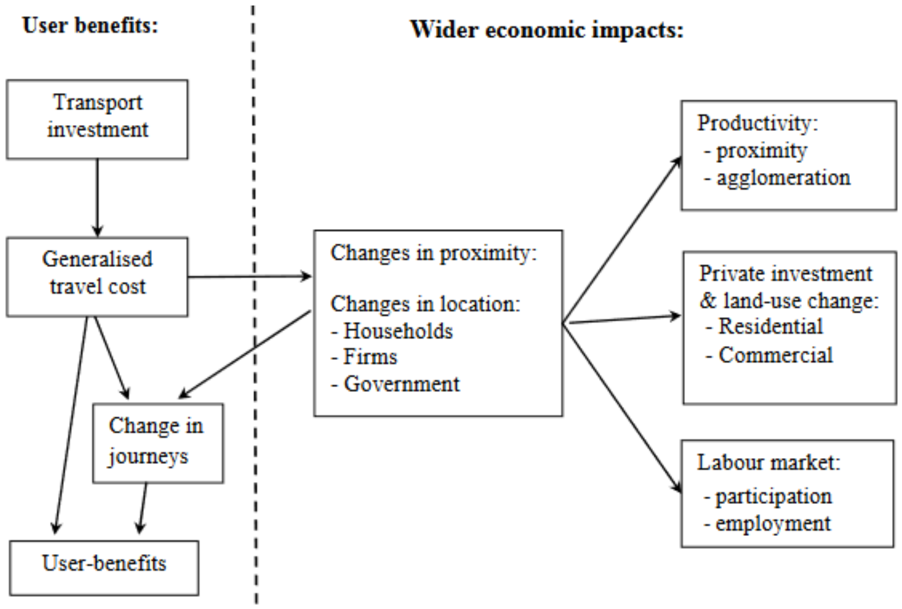


Figure 2.6: The effects of a transport improvement (ITF, 2017).

There are challenges in this approach, related to determining whether a social value above user benefits is achieved, taking into account the wider context in job creation (is there only a displacement or a real change) and related to predicting and quantifying the expected effects.

3 Automated driving functions

This section provides an overview of the ADFs to be tested in L3Pilot, based on the state of the art and an assessment in March 2018. The section provides a guide to the key features of the four major ADFs (traffic jam chauffeur, motorway chauffeur, urban chauffeur and parking chauffeur), and does not focus on the details of design that distinguish between different versions of systems that perform similar driving functions. A more extensive overview of AD function capabilities, limits, and operational design domains (ODD) can be found in Deliverable D4.1.

3.1 Traffic Jam Chauffeur and Motorway Chauffeur

Two SAE Level 3 AD functions are described which will support the driver during motorway driving scenarios: i) **Traffic Jam Chauffeur** and ii) **Motorway Chauffeur**

The Traffic Jam Chauffeur function is available to the driver during low speed driving, including congestion situations, while travelling at speeds up to 65kph in a motorway environment. When activated by the driver, the vehicle will complete the driving task, without the need for driver engagement or supervision. The function will typically provide 10 seconds for the driver to regain control when the vehicle determines that the automated driving can no longer be provided. Traffic Jam Chauffeur is capable of keeping the vehicle in the current lane and maintaining a safe distance to the vehicle in front of the ego vehicle. In some cases, the function will also be able to support lane changes.

The Motorway Chauffeur function extends the speed range of the Traffic Jam Chauffeur function to include speed driving at higher speeds (up to 130kph) in non-congested conditions, on motorways. Additionally, this function will deliver self-deterministic lane changes to maintain progress at the appropriate speed for the road. Like Traffic Jam Chauffeur, the take-over control period is typically 10 seconds. In case of an unresponsive driver, a Minimum Risk Manoeuvre will be performed. This involves stopping the vehicle on the hard shoulder, when it is safe to do so. For the Motorway Chauffeur to perform successfully, two pre-conditions need to be satisfied; there must be a road division between opposite lanes, and the lane/road markers must be visible.

Both functions will work in mainly dry weather conditions (including light rain; excluding snow and fog) and in both day- and night-time.

3.2 Urban Chauffeur

A second function that will be tested during the L3Pilot project is the “Urban Chauffeur”. This function provides conditional automated driving in an urban driving environment. The “urban scenario” can be either a one-way or two-way road. For the “Urban Chauffeur” function to work correctly, it is important that the lane markers on the road are visible and clearly defined.

The vehicles that will be tested during L3Pilot can drive at a speed range between 25 and 50kph. The function is capable of identifying other road users, such as pedestrians and cyclists. Although the tested system will not be able to cope with priority rules (e.g., priority from the right, priority signs), some systems piloted will potentially be able to identify traffic lights and act accordingly. The driver will takeover in unexpected situations (e.g., lack of lane markings, lack of map information, etc.) or in complex scenarios (e.g., high traffic at an intersection). Take-over by the driver will happen when the driver brakes, steers and/or accelerates.

The Urban Chauffeur will perform better during good weather conditions (good light, clear visibility), but it is hoped that will also be able to perform during night conditions. Slippery road, or low visibility due to snow, fog or rain, will limit the performance of the function, and these are instances where the driver is required to takeover control.

3.3 Parking Chauffeur

The third type of AD function tested in L3Pilot is the Parking Chauffeur; an automated parking function. The automated parking functions can handle the actual parking manoeuvre, and the final stages of the drive to the parking space (the distance to the parking space might be limited depending on the function design). These functions typically operate in private or semi-private environments, which includes parking in dedicated spaces or garages. There are certain weather conditions which can limit the operation of the function, for example, snow. Typically, the operation of automated parking functions is limited to low speeds (< 10kph).

The automated operation in parking environments is challenging, since the spatial distance to other objects is small and information typically used by a specific sensor might not be available (e.g., GPS in parking garages). Hence, depending on the implementation of the automated parking function, a training of the parking trajectory may first be required before an actual parking manoeuvre can be performed. Training in this sense means the parking manoeuvre needs to be carried out manually, at least once. During the training run, the functions learn the environment, as well as the relevant positions.

The automated parking functions are designed in a way that during the parking manoeuvre the driver can be either in the vehicle (SAE Level 2), or outside it (SAE Level 4). In cases where the driver is outside the car, they can typically monitor the parking manoeuvre via a separate device (e.g., smartphone).

In the parking environment, functions have to deal with static objects and other slow-moving traffic participants – in particular vulnerable road users (VRU). Automated parking functions need to be capable of detecting and handling these objects and situations in order to park the vehicle. In case a parking function is – for some unknown reason – not capable of handling a situation, the Minimum Risk Manoeuvre of the function will be to brake and bring the vehicle to a standstill and request the driver to take over control. Hence, the automated parking functions are designed either as a Level 2 or as a Level 4 function according to the SAE definition (SAE International, 2016).

4 Background literature on ADF development and testing

This section provides a review of existing literature relating to the development and testing of ADFs, both in simulator and real-world environments. The important issues in ADF design and implementation are highlighted along with remaining challenges for future research.

4.1 The need for ADFs

Vehicle automation is proposed to have a number of benefits, including the potential to: increase the flow and capacity of the road network (Kesting, 2008; Ntousakis, 2015), to contribute to a wide range of economic benefits (Fagnant and Kockelman, 2013), increase shared mobility (Fagnant, 2013), and to reduce energy consumption (Anderson et al., 2014).

Human error is thought to be a contributing factor to over 93% of road accidents (Treat et al., 1979; Sabey and Taylor, 1980). A further assumed benefit of vehicle automation is that, by relieving the human driver of parts of the driving task, human error would be reduced, thus decreasing the number of road traffic accidents. Yet, this potential benefit may not be realised if automation design is not fit for purpose. Human error typically arises out of poor human-system interaction, because of a combination of active failures (e.g., drivers failing to detect a hazard) and latent conditions (e.g., the human-machine interface is designed poorly; Reason, 1990).

Vehicle automation is in its infancy, and it will take some time until a driver is no longer needed under all driving conditions. Until then automated systems will be joint cognitive systems between man and machine (Bibby et al., 1975; Bainbridge, 1983). There will be interactions between driver and system and therefore the possibility of human error, making the human factors of such systems critical to their successful implementation. To realise the full potential of vehicle automation for improving road safety, it is necessary to investigate how automation affects drivers' abilities to interact safely and appropriately with the driving task (Parasuraman et al., 2000; Merat and Lee, 2012).

4.2 Factors affecting the design of a safe and effective Level 3 ADF

There are a number of vehicle/environment-based factors which are known to influence driver performance in the transition from automated to manual control of driving, including the available transition time (time budget), the design of the HMI, and the road and traffic situation.

4.2.1 Time budget

A primary area of interest in the study of transitions in AD in recent years is the time it takes for drivers to resume manual control given a particular time budget or lead time (e.g. Damböck et al., 2012; Gold et al., 2013; Naujoks et al., 2014; Zeeb et al., 2015; Payre et al., 2016). Damböck et al. (2012) was the first to systematically vary the time available to drivers following a take-over request. Comparing take over time of 4 s, 5 s, 6 s, and 8 s, the authors found that, compared to when in manual control, drivers crashed significantly more frequently in all conditions except for the 8 s condition. Gold et al. (2013) examined driver behaviour following an auditory take-over request at either 5 s or 7 s time-to-collisions, with a stationary vehicle in the lane ahead. Results were compared to a baseline group that performed the same task, but in manual driving. The study of Gold et al. (2013) provides a detailed account of behavioural responses to different take-over times (see Figure 4.1). Drivers with a shorter take over times (5 s) were able to react faster in all considered variables, compared to drivers given longer take over times (7 s), but they tended to have fewer glances to the rear and side mirrors before initiating a lane change, and were also less likely to use an indicator. Therefore, drivers who were had a 5 s time take over time showed more erratic behaviour following a take-over.

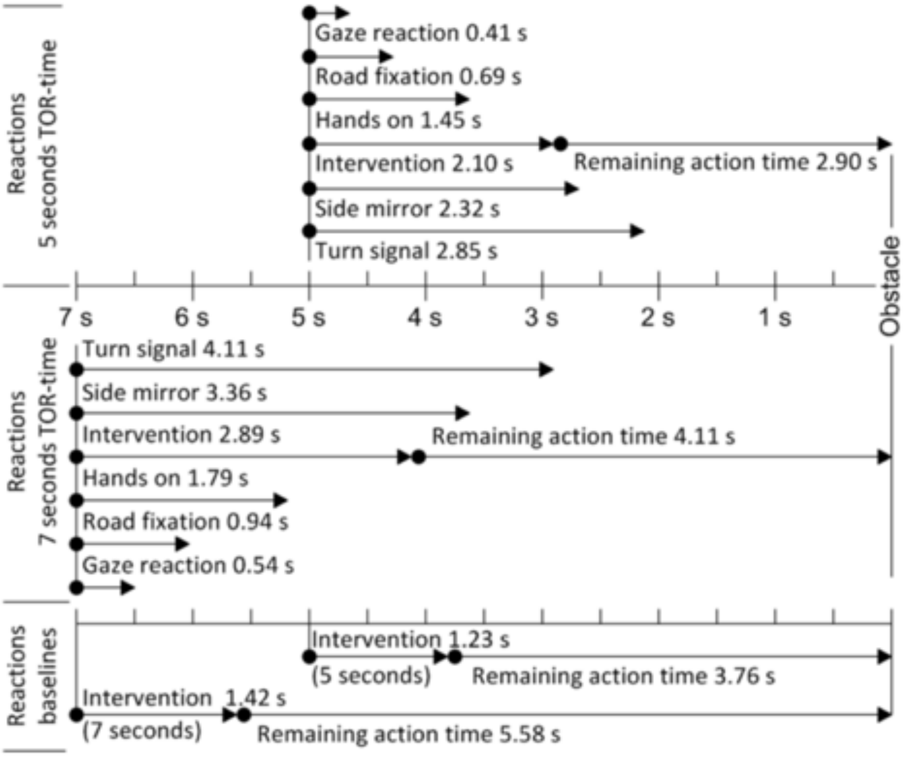


Figure 4.1: Reaction sequences in the Gold et al. (2014) study.

Using a similar rear-end, near-crash, situation, van den Beukel and van der Voort (2013) found that, when given time budgets of 1.5 s and 2.8 s, 47.5% and 12.5% of drivers, respectively, were unable to avoid colliding with a braking lead vehicle. Zeeb et al. (2015) found similar crash rates, where 45% of drivers who were given a time budget of 4.9 s crashed with a lead vehicle, and where 15% of drivers given a 6.6 s time budget, crashed. These studies clearly demonstrate that, in time-pressured take-over scenarios, drivers struggle to resume vehicle control and resolve the situation. However, Louw et al. (2017) showed that a more aggressive response in these situations is not necessarily negative, as these manoeuvres often lead to successful collision avoidances as long as drivers reacted more than 3 s before the eventual collision.

The motivation to focus on driver responses to a resumption of control after different time budgets, arises from a need to define operational and technical parameters for the design of automated driving systems, which, as pointed out by Larsson (2013), provides only a narrow view of how drivers interact with their automated driving system. This is especially true considering that other factors play a part in influencing the success of the transition process. This complicates the understanding of how all the different factors affect driver interaction with the systems, therefore creating a challenge when attempting to design safe and efficient interactions with ADFs.

4.2.2 Human-Machine Interface (HMI)

The Human-Machine Interface (HMI) is used to provide system feedback to users, which is a prerequisite for appropriate human-automation interactions (Norman, 1990). Incorrect or insufficient system feedback results in drivers developing inaccurate mental models, which leads to errors in decision or action (Sarter and Woods, 1995). To ensure that feedback is both correct and sufficient, Norman (1990) proposed a set of four design criteria for automation HMIs, as follows: "Appropriate design should (1) assume the existence of error, (2) it should continually provide feedback, (3) it should continually interact with operators in an effective manner, and (4) it should allow for the worst of situations" (Norman, 1990). Feedback and warning systems of automated vehicles typically communicate information through a range of modalities (for a review, see Manca et al., 2015), in different sequences (Radlmayr et al., 2014), and convey various meanings (Lorenz et al., 2014; Beller et al., 2013).

4.2.3 Scenario

Motorway scenarios have often been used to study driver responses in take-over scenarios in automated driving. Most of these studies have been conducted in driving simulators. In these studies, the scenarios tested typically reflect actual or predicted limitations or boundaries of an automated driving system. These include scenarios where the vehicle is reaching a target destination (e.g. motorway junction), surrounding vehicles or roadworks obstruct intended journey path, technical/sensor failure, and a system limitation in dealing with unaccounted for driving and traffic scenarios, such as an accident or absence of road infrastructure supporting the automation.

Radlmayr et al. (2014), investigated the effect of varying driving situation, and engaging in non-driving related tasks on the process and quality of a system-initiated take-over. The experiment was conducted in a high-fidelity driving simulator (e.g. immersive visual experience, motion-based simulator), using the standardised visual Surrogate Reference Task (SuRT) (a standardised visual distraction task), and the cognitive n-back task (a task involving recollection of numbers presented n positions back in an ongoing auditory list) to simulate the non-driving related tasks. The study included four driving situations, each with a time budget of 7 s: In Situation 1, an obstacle appeared in the middle lane, while the right and left lane were blocked by vehicles at the time of the take-over request (TOR). In Situation 2, an obstacle appeared in the right lane while no other vehicles were present during the situation. In Situation 3, the obstacle appeared on the left lane. Situation 4 closely resembled Situation 1, except that the two adjacent lanes were not blocked. The authors used different dependent variables to infer the quality of the transition. These included take-over time, longitudinal acceleration, time to collision (TTC), the total number of collisions during take-over, results from a Detection Response Task (DRT) and the subjective rating. The authors concluded that traffic scenario and traffic density had a substantial effect on take-over quality in a motorway setting. Similarly, Kircher et al. (2014) found that drivers' response times were moderated by whether the driver was pre-warned, and by the type of scenario. Traffic density (a measure of traffic scenario complexity) has also been shown to influence how long drivers need to regain situation awareness and resume control (Jamson et al., 2013; Gold et al., 2016), although Naujoks et al. (2014), found no effect of traffic scenarios in a take-over event.

Automated vehicles on motorways also have an impact on the driving behaviour of others. A driver simulator study of Gouy et al. (2014) showed an effect on driver's time headway (THW) to a lead vehicle when driving next to a platoon of automated trucks with different time headways in a motorway environment. The drivers adapted their behaviour towards the platoon to some extent by driving with shorter THWs when exposed to a platoon with shorter THWs and also spending more time below the critical threshold of 1s. This increases the chance of a collision.

A number of inter-related psychological factors that have been linked to drivers' capacity to interact safely with automated driving systems, including driver trust (Lee and See, 2004; Hergeth et al., 2016), locus of control (Stanton and Young, 2005), complacency (Bagheri and Jamieson, 2004; Parasuraman and Manzey, 2010), mental models (Moray, 1990; Sarter et al., 2007; Flemisch et al., 2012), driver state (Rauch et al., 2009; Neubauer et al., 2012; Jamson et al., 2013; Louw & Merat, 2017), mental workload (MWL; De Waard, 1996; Collet et al., 2003), and Situation Awareness (SA; Endsley, 1995a; Merat and Jamson, 2009; Kircher et al., 2014).

Of these factors Situation Awareness (SA) is considered one of the most important factors that is predictive of performance and safety (Parasuraman et al., 2008). Endsley (1988) defines SA as, *"the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future".* Simply put, good SA involves being aware of various elements of information in the environment, relevant to successful task performance (Situation Awareness Level 1, Perception), being able to interpret the meaning, context and significance of that information (SA Level 2, Comprehension), and being able to predict the future state or configuration of the environmental conditions (SA Level 3, Projection; Endsley, 1995a; Figure 4.2).

In a critical review of the literature, de Winter et al. (2014), argued that SA can be affected by automated driving (AD). Studies of drivers' eye movements, tests of object detection, engaging in tasks unrelated to driving, and responses to critical incidents can be used to measure the effects of AD on SA.

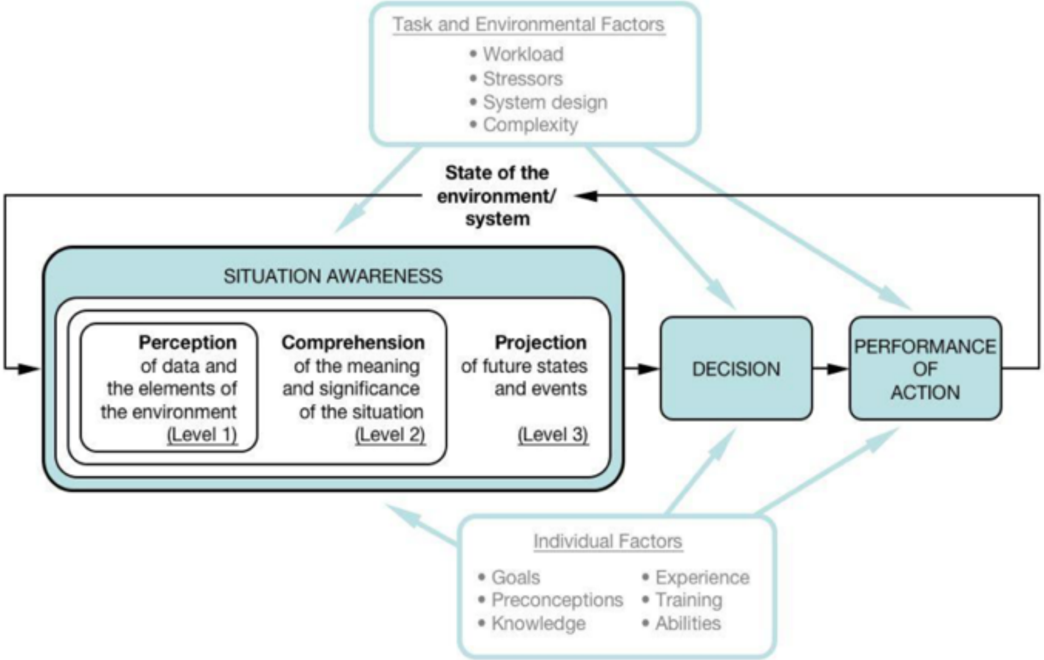


Figure 4.2: Endsley's model of SA. Adapted from Endsley (1995a), and Endsley et al. (2000), in Jin (2008).

In terms of drivers' eye movements, a few studies (e.g., Carsten et al., 2012; Damböck et al., 2013) have found that, compared to manual driving, drivers in AD were overwhelmingly less likely to monitor the road, especially the centre of the visual field. Carsten et al. (2012), for example, found that, for drivers who did not focus on the DVD player, which was operating at SAE Level 3 (Conditional Automation), their eyes were focussed on the central region of the road only 53% of the time, compared to 72% in manual driving. While these results do not allow us to draw conclusions on the level of detail of SA, they do give some indication that drivers under automation are more inclined to lose SA for the driving task/environment during use of a SAE Level 3 ADF, mainly because they are not monitoring the driving task. Moreover, it is well established that the more visual attention is paid to the road scene the better is the level of SA (Green, 1999; Chaparro et al., 1999). Therefore, it follows that the reverse would also be true. However, not many studies have examined eye gaze behaviour for drivers with varying levels of engagement with the driving task during Level 3 automation, which is relevant given the different tasks/activities that drivers could engage with during automation. One example is that of Merat et al. (2014), who analysed drivers' reaction times after automation disengaged automatically, if they were not looking at the road centre for longer than 10 s. The authors found that there was a 10-15 s lag time between when automation disengaged and when drivers resumed control, based on their visual attention back to the road centre. In another driving simulator study, Gold et al. (2013) showed that, in-between a take-over request and resumption of control, drivers tended to fixate on side mirrors, which suggests that this time is used to regain SA before resuming control. Lorenz et al. (2014) reported that drivers are quick to direct visual attention to the road scene, but struggle to instantly understand the situation. However, the precise eye-movement patterns underlying the process of building up SA during take-over situations is not currently clear (although this is also true to some extent for manual driving).

There is a consensus in the literature that AD encourages non-driving related activities, compared to manual driving. It is an important reason for the development of AD. Carsten et al. (2012) report that, during Level 3 ADF operation, drivers were more likely to use a DVD player than during manual driving (32.5% vs 2.6% of the time) and were also more likely to use the radio (54.1% vs. 41.4% of the time) and even read a magazine (9% of the time). This is confirmed by other studies, with Llaneras et al. (2013) finding that, during Level 3 ADF operation, half of the participants texted or emailed during a 1.5h-2h drive. The authors found that participants engaged in a number of other tasks during AD, including eating, reaching for an item in the rear compartment, dialling and talking on the cell phone. The fact that drivers chose to engage in non-driving related tasks does not necessarily reflect the extent to which SA has been lost, or indeed that they have impaired SA at all. Similar to measures of eye movements, it gives an indication of the extent to which SA could possibly be lost. It seems reasonable to assume that a lack of attention to the main driving task is likely to disengage drivers, reducing their ability to react promptly, and suitably, during failures of the automated system. However, a limitation of this method is that the authors do not link the type of reaction to any safety or performance-related outcomes, which would be useful to form a holistic view of the importance of SA to performance in the transition.

Measuring drivers' reactions to critical events while in automated driving mode has become a standard methodology for the assessment of the impact of AD on performance. As part of the CityMobil project, Merat and Jamson (2008), examined driver reactions to a number of 'critical' incidents, including merger from the left, an oncoming car turning across the vehicle path, the presence of traffic lights, and the presence of a parked car. Drivers' responses in manual driving were compared to those in AD, where a critical incident triggered automation disengagement and the driver was expected to respond. The authors measured 'anticipation', defined as drivers' ability to predict and understand the behaviour of traffic during these critical events, to conclude on drivers' SA. Anticipation was measured as the difference in time between the lead car's brake lights coming into sight, and when drivers pressed their brake pedal. Results showed that, for the three critical events, drivers in the automated driving condition braked on average 1.5 s later than in the manual driving condition. In the critical event with the parked car, 28 of the 38 drivers in the automated driving condition braked after the collision warning alarm was emitted. One possible explanation of this high proportion of late braking is that drivers were less aware of the unfolding events, but, the authors suggest it may also be because drivers over-relied on the automated system, concluding that the use of an automated driving system may reduce drivers' situational awareness and may also cause drivers to become complacent behind the wheel.

This short section has provided an overview of the likely consequences of AD engagement on driver behaviour and response during critical situations. Most of the studies reported are based on data collected in driving simulators, therefore, it will be interesting to investigate how and whether this behaviour is affected by AD engagement in the real-world.

4.3 Challenges for ADF implementation – real-world examples

Urban driving is more complex than motorway driving, since urban drivers typically encounter many different types of vehicles and road users (e.g., trucks, buses, motorcycles, scooters, pedestrian, cyclists, and skaters). Urban driving is also characterized by many complex driving manoeuvres, such as handling of intersections, and interacting with different types of road users at the same time. In contrast, motorway driving is largely composed of car-following and lane changing manoeuvres.

Urban driving is also less predictable than motorway driving. For example, on motorways, a lane change manoeuvre of another vehicle can be easier to anticipate than in an urban area. In the urban area, higher traffic density and variation in lane width and the frequency of obstructions in lane means that surrounding vehicle speed and lane position are not as effective predictors of upcoming lane changes as they are in the motorway environment. Another challenge in urban areas is the presence of pedestrians, who have more degrees of freedom in their movement choices than other vehicles or cyclists, meaning that it is difficult to predict their upcoming manoeuvre. For example, in order to come to a complete stop, motorists and cyclists need a certain braking distances while pedestrians need almost no braking distance. Considering a U-turn, pedestrians can change their direction on the spot while motorists and cyclists need space and time to complete such a manoeuvre. This features of urban road user interactions demonstrate that the development of an urban ADF is more challenging than a motorway ADF. Motorways are structured environments, whereas urban areas are rather unstructured. This poses challenges to the algorithms underlying ADFs. Hence, the algorithms for situation interpretation, decision making and trajectory planning have to be different in these two environments.

4.3.1 Past studies on Urban Chauffeur ADFs – challenges in implementation

Nevertheless, over the last two decades, technological advances have greatly improved and significant effort has been put into the development of urban ADFs. In 2007, the first urban challenge was funded by the US Department of Defence (Defence Advanced Research Projects Agency, DARPA). The DARPA urban challenge (DUC) was conducted in a mock-up urban environment on a closed airfield. This challenge included emergency stopping, interacting with other motorized traffic, handling intersections, and following other vehicles.

The interaction with vulnerable road users (VRU) and traffic lights were not included in this field test. The project identified several challenges of urban ADF operation that needed to be addressed, mostly falling into the Technical and Traffic Evaluation area of the L3Pilot project.

From the Technical and Traffic standpoint, there were observed system processing delays resulting from the high number of sensors for detection and navigation that were installed on the test vehicles (Behringer, 2004). System capabilities were also a topic recommended for future work. Specifically, to create a safe interaction between the vehicle and other road users, the urban ADF needs to be able to anticipate the behaviour of other road users and classify that behaviour into good or bad performance in order to adapt its own driving. The function also needs to capture non-verbal communication clues and learn from mistakes that it makes. In addition, it needed to trust other road users, and be able to cooperate with them. These issues all fall under the general challenge of evaluating the ADF (Campbell, 2010).

These technical challenges of urban ADFs have been addressed across a number of research projects in the past decade, including the Intelligent Vehicle Future Challenge in China (e.g. Broggi et al., 2105; Sun et al., 2014; Mei et al., 2012), the Grand Cooperative Driving Challenge (GCDC) and adaptive ((Rodarius et al., 2015). In the latter case, a vehicle platooning scenario was introduced, whereby a platoon had to start at a red light and join another platoon ahead. Platooning performance and shockwave damping were evaluated (van Nunen, 2012). This platooning function has since become a pre-requisite for urban ADFs in the second-round of the GCDC i.e. vehicles should be able to merge from one platoon into another in response to a lane closure. The other urban ADF task was an intersection manoeuvre in which three cars approached a T-intersection, where they were required to cross the intersection simultaneously without stopping and/or colliding (Dolk, 2017).

In 2013, Broggi et al. (2015) performed an automated driving experiment in the city of Parma, Italy. In this experiment, 1.7 km were driven in urban areas and the tasks included waypoint following, lane following, and handling a roundabout, traffic lights, and T-intersections. The experiment could be completed without human interference. Also in 2013 a drive from Mannheim to Pforzheim (103 km) including 54 km urban driving was completed by an automated vehicle. The automated vehicle had to handle 155 traffic lights as well as 18 intersections in various traffic conditions. It also had to react to parked cars, preceding and oncoming traffic, bicyclists and pedestrians. The ride was completed without any sudden intervention of the safety driver (Ziegler, 2014).

These challenges and experiments show that the development of urban ADFs has taken significant steps since the first urban challenge in 2007, and progress continues to be made in current research projects. For example, the European research project InterAct is currently analysing human-human interaction strategies in order to implement cooperative and intuitive interactions between AVs and other road users (see www.interact-roadautomation.eu). The German project PEGASUS (see www.pegasus.projekt.info) and the European project

ENABLE-S3 (www.enable-s3.eu) address issues with regard to verification and validation of automated driving functions.

In addition, efforts have been made to investigate the effects of automation on driving performance. An interesting field of research is take-over requests and the (safety) effects of resuming control over the vehicle. Another topic of research is skill loss due to automated driving (e.g., Tofetti et al., 2009). In addition, simulation studies have been administered testing the effects of GLOSA (green light optimization speed advisory) and platooning on traffic flow and environment in urban areas (e.g., Lionis, 2016, Krajzewicz et al., 2012; van Arem et al., 2006).

4.3.2 Ongoing studies (disengagements trends in US)

California, Florida, Nevada, and Michigan, USA allow unmanned driving on public roads. California established the Autonomous Vehicle Tester Program in 2012 regulating automated driving on public roads (DMV California, 2018). Automated cars are now legally allowed to drive driverless without steering wheel, foot pedals, and mirrors within the state of California. The revised regulation governs the safe introduction of automated vehicles onto public roads (Hawkins, 2017). Car manufacturers can apply to the program in order to test their automated vehicles on public roads. After being accepted to the program, manufacturers are required to report every crash involving an automated vehicle. Once a year, manufacturers also need to submit an extensive report of all disengagements (DMV California, 2018). These reports can be used to gain first insight into the effects of an automated vehicle on safety. The vast majority of automated driving is done in (sub-) urban areas.

Between October 2014 and December 2017, 52 crashes involving an automated vehicle have been reported, although some of these involved the vehicle when it was in manual driving mode. Dotzauer et al. (2017) analysed reports of crashes that occurred until December 2016. The smallest proportion of the crashes were crashes where the automated vehicles was considered to be at-fault. Half of the reported crashes were rear-end collision caused by another driver and most of those occurred while the automated vehicle tried to merge into traffic. California crash data of human drivers indicates that about 20-30% of all crashes are rear end collisions. It may be hypothesized that the behaviour displayed by the automated vehicle could not be anticipated by the following driver accurately. The gap for merging may have been assessed as sufficiently large enough by the driver, but not by the automated vehicle. A further example crash type involving the automated vehicles is where another driver is leaving their lane. It appears that the automated vehicle has limited capabilities to assess and anticipate the behaviour of other road users and then act upon this assessment (i.e. adjusting position in its lane in order to avoid a collision).

The 1183 disengagement reports – whether driver- (687) or system-initiated (496) – were also analysed. (DMV California, 2018). The main reason for driver-initiated disengagements was because drivers did not feel comfortable being driven by an automated vehicle in the prevailing situation. 50% of system-initiated disengagements took place because of system failures, for example, crashes of the computer, localization issues, insufficient calculation power, or transmission of CAN Bus data (Dotzauer et al., 2017). The other half occurred because the automation reached predefined limits, including construction areas, missing lane marking, emergency vehicles, or traffic lights. In addition, limits with regard to other road users are also predefined. The automated vehicle also disengages when traffic density is high during a lane change manoeuvre, it detects unexpected behaviour of another motorist, density of pedestrians is high, or another motorist drives recklessly (Dotzauer et al., 2017). Many of these systems limitations match with those identified by Campbell et al. (2010). This demonstrates the wide range of challenges that an effective urban ADF must overcome to perform safely in this complex environment surrounded by unpredictable road user behaviours whilst being accepted by the user.

It should however be noted that these studies in California are limited to some extent by the type of driver inside the automated vehicle. Until 2017, safety drivers were required in all automated vehicles and needed to be attentive at all times (i.e., drivers were not allowed to engage in non-driving-related secondary tasks). Nonetheless, because of the crash and disengagement reports, we are able to gain first knowledge of the effects of automated vehicles in mixed traffic conditions in urban areas and identify areas of interest to investigate further.

5 Methodology: Research question generation

This chapter describes the process by which the appropriate Research Questions for this pilot study have been generated and selected, and how associated hypotheses have been defined. It then goes on to describe how Research Questions have been used to identify the logging needs required for the evaluation activities of the project.

5.1 Top-down approach: From evaluation and impact areas to RQs

The L3Pilot approach uses the FESTA V methodology, but introduces some adaptations to account for the ongoing technical validation of the AD functions for test in this project, and the subsequent likelihood that there could be changes in characteristics of the function as the project progresses.

An early project review of the AD functions that are likely to be available for test was conducted to assist with the task of setting the Research Questions **(FESTA Step 1: Selection and description of functions)**. This revealed three AD functions: Motorway Chauffeur, Urban Chauffeur and Parking Chauffeur. These functions were used as a guide for research question generation but the fluidity of this list, or at least the fluidity of the precise technical capabilities of the systems were acknowledged during the process. For this reason, the definition of specific use cases and situations **(FESTA Step 2)** was postponed until the specific details of the AD functions and the details of the field experiments could be confirmed. Instead, the focus was on developing high-level Research Questions to meet the objectives of the project **(FESTA Step 3: Identification of the Research Questions)**. The top-down approach used for that was the Impact area approach recommended by FESTA Handbook. The basic principle for generating Research Questions and hypotheses using this top-down approach lies in a theoretical understanding of the factors that influence the different impact areas.

The research question setting process acknowledged the considerable range of areas in which the AD functions could have possible impacts, including:

- User preferences and reactions, and Human Machine Interaction (HMI);
- Driving behaviour in terms of system usage and behavioural adaptation;
- Safety impacts deduced from critical behaviours such as near-crashes and speed behaviour;
- Impacts on the traffic flow and non-automated road users including vulnerable road users;
- Impacts on environment and efficiency;
- Societal and socio-economic projections.
- Deployment potential.

The generation of the Research Questions was a top-down process, structured around the four evaluation areas of L3Pilot: Technical & Traffic Evaluation, User and Acceptance Evaluation, Impact Evaluation, and Socio-Economic Evaluation. The use of a categorisation structure that would be employed later in the project at the Evaluation stage ensures consistency in the preparation and evaluation frameworks and in the terminology used throughout the project.

The observation of user behaviour in both automated and manual driving modes is expected to create impacts in eight areas; Safety, Mobility, Efficiency, Environment, Acceptance & Awareness, User Experience, System Performance, and Socio-economic. These impact areas are used to create a two-dimensional research question framework (or matrix) in which each research question is allocated to a specific Evaluation area/Impact area combination (Table 5.1). The use of a two-dimensional research question framework allows the researcher to ensure that a research question is contributing to a specific evaluation area, and under which impact area it is doing so. The allocation of a research question to a specific cell (or cells) of the matrix also allows for identification of overlap and thus allows for more efficient division of workload amongst analysts at the evaluation stage.

Table 5.1: Research question setting framework

		Evaluation Area			
		1. Technical & Traffic Evaluation	2. User & Acceptance Evaluation	3. Impact Evaluation	4. Socio-Economic Impact Evaluation
Impact Area	a. Safety				
	b. Mobility				
	c. Efficiency				
	d. Environment				
	e. Acceptance & awareness				
	f. User experience				
	g. System performance				
	h. Socio-economic				

The research question setting process was started by a thorough review of the existing literature to identify main factors related to different impact areas and knowledge gaps relating to user or driving behaviours when using automated driving functions. In addition, repeated cycles of 'research question generation – review – edit and addition' were conducted to ensure coverage of all major topics with the potential to affect future knowledge, society or business.

The first stage of research question setting focussed on high-level questions that needed to be answered per evaluation area. These are displayed below, categorised per evaluation area **(RQ Level 1)** (Table 5.2).

Table 5.2: Research Question Level 1

Technical and Traffic Evaluation
• What is the system's technical performance? • What is the impact on the driving behaviour? (Considered during manual driving and AD) • What is the impact of ADF on the interaction with other road users? • What is the impact of the ADF on the behaviour of other traffic participants?
These questions focus on the readiness of the ADF for implementation and on impacts of the ADF on driver behaviours, such as speed and headway distribution, and interaction with other road users.
User and Acceptance Evaluation
• What is the impact on user acceptance and awareness? • What is the user experience?
These questions focus on facets of the user experienced including use, transfer of control, interaction.
• What is the impact on safety? • What is impact of ADF on environmental aspects? • What is the impact of ADF on travel behaviour? (Exposure)
These questions focus on the potential wider impacts of ADFs on safety, environment and mobility.
Socio-Economic Impact Evaluation
• What is the socio-economic impacts of ADF?
These questions focus on the scaling up of the impacts of the ADFs to the Europe level, and the cost-benefit analysis of their use on European roads.

The second stage of research question setting involved development of these into more detailed questions relating to specific components of the higher level questions (e.g. different aspects of user acceptance **(RQ Level 2)**, before going one step further to ask questions about specific use cases, driving situations or performance indicators **(RQ Level 3)**, where appropriate**.** This led to the generation of 75 distinct questions at RQ Level 3. An example for one question at RQ Level 1 is shown below (see Table 5.3). Throughout the research question generation process, each RQ was assigned to an Evaluation area, Impact area, and AD function, to ensure that each RQ was contributing to a specific project objective.

Table 5.3: An example of the three-level approach to setting Research Questions

RQ Level 1	RQ Level 2	RQ Level 3
What is the impact on user acceptance & awareness?	Are drivers willing to use an ADF?	Are drivers willing to use an ADF?
	How much are drivers willing to pay for the ADF?	How much are drivers willing to pay for the ADF?
	What is the user acceptance of the ADF?	What is the perceived safety of the ADF?
		What is the perceived comfort of the ADF?
		What is the perceived reliability of the ADF?
		What is the perceived usefulness of the ADF?
		What is the perceived trust of the ADF?
		How is user acceptance influenced by system behaviour in unexpected use cases?
	What is the impact of the ADF on driver state?	What is the impact of ADF on driver stress?
		What is the impact of ADF on driver fatigue?
		What is the impact of ADF on driver workload?
	What is the impact of ADF use on driver awareness?	What is the effect of ADF use on driver attention to the road/other road users?
		What is the impact of ADF on driver risk perception?
	What are drivers' expectations regarding system features?	What are drivers' expectations regarding system features?

5.2 Hypothesis generation

From the more detailed research question (RQ Level 3), specific hypotheses were generated **(FESTA Step 4: Creation of hypotheses)** to guide the 'Evaluate' stage of the FESTA methodology. This process led to the generation of greater than 100 detailed hypotheses. The output of this process was a spreadsheet of Research Questions and associated hypotheses defined by Evaluation area, Impact area, and AD function. This process has the effect of increasing the depth of questions asked within the project, and linking hypotheses to specific ADFs and use cases.

5.3 Identifying data logging needs

In order to test the hypotheses, logging needs for objective data have to be determined for the pilot studies. Section 5.3 presents the methodology for identifying logging needs, then the logging needs for the pilot are presented and linked to the Research Questions (and their associated hypotheses) in Section 5.4.

As prescribed by the FESTA Handbook, the next step after hypothesis generation is to derive the logging needs for each hypothesis. For this process, a distinction is made between subjective and objective data, which have to be collected during the pilot. Whereas subjective data will be collected by questionnaires that are evaluated in the User and Acceptance Evaluation, objective data will be extracted mostly from the data logging systems in the vehicle, from additional cameras mounted on the vehicle, and where required, from external data sources as well (e.g. weather information, road type etc.).

This data is afterwards analysed in the Technical and Traffic Evaluation and the User and Acceptance Evaluation (where appropriate). A list of signals for logging in the vehicle will be derived from the hypotheses defined previously, (see Table 5.4). For the evaluation areas the Research Questions are specified on three levels. Based on these the hypotheses are defined. In this context, the indicators necessary for answering the hypotheses are identified. Finally, the logging needs in terms of signals to be recorded are derived based on the formulas for calculating the indicators.

Table 5.4: An example of how logging requirements were defined per hypothesis

Evaluation area	Technical & traffic
RQ level 1	What is the impact of the ADF on driving behaviour?
RQ level 2	What is the impact of the ADF on driven speed in different scenarios?
RQ level 3	What is the impact of the ADF on driven speed in driving scenario X?
Hypothesis	e.g. 1: There is no difference in the driven mean speed for the ADF compared to manual driving. e.g. 2: There is no difference in the standard deviation of speed for the ADF compared to manual driving.

Evaluation area	Technical & traffic
Performance indicators required	Mean speed, standard deviation of speed, max speed, plot (speed/time)
Logging requirements / sensors available	CAN bus of vehicle: Ego speed in x-direction

In order to establish a structured process for deriving the logging needs, a table containing all Research Questions and their associated hypotheses for the respective evaluation areas (Technical and Traffic, User and Acceptance, Impact and Socio-economic Impact Evaluation) was prepared. Within this table, all Research Questions have an individual RQ-ID. The preliminary hypotheses derived for each RQ are also ascribed the same RQ-ID. As shown in Table 5.5, the logging needs are added to the table by linking them in a matrix structure with hypotheses, as per the FESTA 'V' approach. With this structure, all relevant logging needs, e.g. lateral acceleration, could be linked to a hypotheses, and thus to a Research Question.

For quantitative analyses, the hypotheses will often include the required performance indicator in its phrasing. In this case, the selection of the required performance indicator and desired logging needs is rather simple. In other cases, the hypotheses are less prescriptive in terms of the required performance indicator, and thus surrogate measures need to be identified from past research. Following the definition of these surrogate measures, the logging needs for these hypotheses can be defined. In the latter example, it is often necessary to define and measure a new performance indicator, especially in situations where the research is novel i.e. the first on-road test of a L3 ADF.

Table 5.5: Framework for allocation of logging requirements to research questions

						Logging needs			
						Ego vehicle data			..
Evaluation area	RQ-ID	RQ Level 1	RQ Level 2	Hypo-theses	Performance indicators	Throttle position	Brake pressure	Longitudinal velocity	…
Technical & traffic	RQ-T1					0	0	0	1
	RQ-T2					0	0	1	1
	RQ-T3					0	0	1	0
	RQ-T4					0	0	1	1
	RQ-T5					0	0	1	1
	…					0	0	1	1

This process of deriving logging needs has been carried out for all defined hypotheses. In this deliverable, logging needs are linked to their higher level research questions (RQ Level 2) for ease of presentation. A full summary of required logging needs per hypothesis is provided in D3.4 (Final Evaluation Plan). However, the stepwise progression from Research Questions to hypotheses to performance indicators, then logging needs has been followed in this work.

5.4 List of logging needs for L3Pilot

The logging needs were derived by using the method described in Section 5.3. In order to structure these, data categories were introduced. These 'types of data' with their description are introduced in Table 5.6.

Table 5.6: Types of data/logging needs and their description

Type of data	Description
General information on vehicle	General information on attributes of the vehicle, e.g., wheelbase, length and width of vehicle. No continuous signals.
Ego vehicle data	All signals, which give information on the current state of the ego vehicle, e.g., speed of ego vehicle in x-direction, steering wheel angle.
Information on system	Here, all signals giving information about the state of the evaluated ADF, e.g. whether the ADF is active.
Object data, information for relevant objects	All signals referring to relevant dynamic and static objects, e.g., distance to relevant object in x-direction
Lane marking data	In order to detect if a vehicle enters the lane of the ego-vehicle, information on lane markings is necessary, e.g., distance to left lane marking
Other environmental information (e.g. from map)	In order to know the number of lanes or the road type the vehicle is driving on, additional data is necessary.
Traffic sign data	Here, all signals giving information on traffic signs are represented, e.g., speed limit of current road section.
Video	All video signals for investigating relevant situations in detail, e.g., front video data

The project requires a series of different data streams to test the diverse range of hypotheses across the four evaluation areas. This necessitates a range of sensors with the ability to measure a vast array of driver, vehicle and environmental factors.

In Table 5.7 the Research Questions are linked to the types of logging needs. Here, the Research Questions are only linked to the groups of logging needs and not to the single signals. However, a detailed overview for each signal will be reported in D3.4, following precise specification of the data logging capabilities at each pilot site in SP5.

Table 5.7: Types of data/logging needs linked to Research Questions

Type of data	Research question
General information on vehicle	RQ-T1, RQ-T2, RQ-T3, RQ-T4, RQ-T5, RQ-T6, RQ-T8, RQ-T14, RQ-T15, RQ-T16, RQ-T17, RQ-T18, , RQ-I1, RQ-I2, RQ-I3, RQ-I4, RQ-I5, RQ-I6, RQ-I7, RQ-I8
Ego vehicle data	RQ-T1, RQ-T2, RQ-T4, RQ-T5, RQ-T6, RQ-T7, RQ-T8, RQ-T9, RQ-T10, RQ-T11, RQ-T12, RQ-T13, RQ-T14, RQ-T15, RQ-T16, RQ-T17, RQ-T18, , RQ-I1, RQ-I2, RQ-I3, RQ-I4, RQ-I5, RQ-I6, RQ-I7, RQ-I8
Information on system	RQ-T1, RQ-T2, RQ-T3, RQ-T4, RQ-T5, RQ-T6, RQ-T8, RQ-T14, RQ-T15, RQ-T16, RQ-T17, RQ-T18, RQ-U10, RQ-U11, RQ-I1, RQ-I2, RQ-I3, RQ-I4, RQ-I5, RQ-I6, RQ-I7, RQ-I8
Object data, information for relevant objects	RQ-T1, RQ-T2, RQ-T4, RQ-T7, RQ-T10, RQ-T11, RQ-T14, RQ-T15, RQ-T16, RQ-T17, RQ-T18 RQ-I1, RQ-I2, RQ-I3, RQ-I4, RQ-I5, RQ-I6, RQ-I7, RQ-I8
Lane marking data	RQ-T1, RQ-T2, RQ-T4, RQ-T7, RQ-T10, RQ-T11, RQ-T14, RQ-T15, RQ-T16, RQ-T17, RQ-T18 RQ-I1, RQ-I2, RQ-I3, RQ-I4, RQ-I5, RQ-I6, RQ-I7, RQ-I8
Other environmental information (e.g. from map)	RQ-T1, RQ-T2, RQ-T4, RQ-T7, RQ-T10, RQ-T11, RQ-T14, RQ-T15, RQ-T16, RQ-T17, RQ-T18 RQ-I1, RQ-I2, RQ-I3, RQ-I4, RQ-I5, RQ-I6, RQ-I7, RQ-I8
Video	RQ-T1, RQ-T2, RQ-T3, RQ-T5, RQ-T11, RQ-T12, RQ-T14, RQ-T16, RQ-T17, RQ-T18, RQ-U6, RQ-U9, RQ-U10

For the logging of video data streams, a more detailed overview is given in Table 5.8. In this context, it is necessary to distinguish between video data facing in front in the direction of the road ahead, video data facing the rear direction and video facing to the left and right side of the vehicle. Additionally video data facing inside the vehicle for monitoring the middle console and for monitoring body and face of the driver is mentioned. It is important to be aware of the ethical issues of recording in-vehicle video, especially featuring the driver. As a result, only those video views that are required for annotation for a specific performance indicator will be recorded.

Table 5.8: Types of video data for logging

Type of video data	Research question
Video front	RQ-T1, RQ-T2, RQ-T3, RQ-T7, RQ-T10, RQ-T11, RQ-T12, RQ-T14, RQ-T17
Video rear	RQ-T1, RQ-T2, RQ-T3, RQ-T7, RQ-T10, RQ-T11, RQ-T12, RQ-T14, RQ-T15, RQ-T16, RQ-T17
Video side left & right	RQ-T1, RQ-T2, RQ-T3, RQ-T7, RQ-T10, RQ-T11, RQ-T12, RQ-T14, RQ-T17
Video inside (middle console, HMI)	RQ-U10
Video driver face/ body	RQ-U9, RQ-U10
Video driver hands	RQ-U10
Video driver feet	RQ-U10

The user and acceptance evaluation of the project will also require input from subjective data sources. These sources are highly pertinent given that this work involves one of the first on-road testing of SAE Level 3 AD functions in Europe. To this end, both an annual survey of the public and specific pilot site questions posed to the test participant drivers of the pilot vehicles will be developed to tackle issues such as system acceptance, perceived trust, perceived safety, willingness to pay, driver experience of take over scenarios, and many more. The goal of the subjective data collection is to gain an insight into user experience above and beyond what can be observed in the objective AD function testing. It also allows surveying of a significantly greater sample size than can be achieved through only the participants of the pilot tests.

5.5 Bottom up approach: from logging needs to RQs and hypotheses

The top-down approach to setting Research Questions was supported by a bottom-up check at a later stage, whereby in line with the FESTA methodology the developed hypotheses and research questions were cross-checked for feasibility based on the data logging capabilities within the project. This was achieved through workshop-based exercises between research institutions and vehicle owners. This was then fed into a process of research question prioritisation based on factors including feasibility of answering the question in a pilot study, available data logging capabilities and test sites characteristics, data processing and coding demands, data management and ethical constraints, project resources and timescale, and research importance. A scoring system was employed to aid with the prioritisation process.

The output from the top-down and bottom-up approaches is a refined list of Research Questions that not only provide the required outputs from the piloting work in each Evaluation area, but do so in a structured and resource efficient manner. Later deliverables will detail how requirements on experimental procedures are set for each Research Questions (D3.2) and how evaluation methods are allocated for analysis and evaluation of them (D3.3) so as to ensure valid data collection and streamline the evaluation process. As ADFs and test plans are still under development, the setting of methodology be tied in with this stage. Therefore, the final version of the methodology will be reported in the Final Evaluation Plan (D3.4) where some changes also to the topics reported in this deliverable are likely.

6 Summary of Research Questions and associated logging needs

This section lists Research Questions per Evaluation area. The top two levels of research question are displayed, with their associated logging needs. In all cases, research questions have been further developed to RQ Level 3 and the hypothesis level, and linked to an Impact area under which their effects will be evaluated.

Note that where data exists for a meaningful analysis of between-subjects factors such as age, gender, personality, and driving experience, these covariates will be considered. In addition, where the experimental design allows, possible longitudinal effects of system interaction will be considered. For this reason, the effects of these factors are not listed as separate Research Questions in this chapter.

Additional detail relating to hypotheses for each RQ can be found by referring to the Final Evaluation Plan (D3.4).

Note that where vehicle data is listed as the logging requirement, this includes alternative external data sources in instances where the required data is not available from the vehicle. For example, if weather information cannot be extracted from vehicle sensors and video annotation, external sources of this data will be sought.

6.1 Technical & Traffic Evaluation

RQ-ID	RQ Level 1	RQ Level 2	Logging requirements	Performance Indicators
RQ-T1	What is the system's technical performance?	How reliable is system performance in a given driving and traffic scenario?	Vehicle data	System status, distribution of longitudinal velocity per driving scenario and situation variables
RQ-T2		How often and under which circumstances do the ADFs issue a take-over request?	Vehicle data	System status, distribution of longitudinal velocity per driving scenario and situation variables
RQ-T3			Vehicle data	System status, distribution of longitudinal velocity per driving scenario and situation variables
RQ-T4	What is the impact on the own driving behaviour?	Are there any traffic violations while using the ADF?	Vehicle data	Distribution of difference between speed and speed limit, distribution of distances to other objects, frequency of overtaking manoeuvres in overtaking prohibitions
RQ-T5		How do take-over requests affect driving?	Vehicle data	Distribution of lateral and longitudinal acceleration and velocity
RQ-T6		What is the impact of ADF on the driving comfort?	Vehicle data	Distribution of longitudinal acceleration
			Vehicle data	Distribution of lateral acceleration
RQ-T7		What is the impact of ADF on the accuracy of driving?	Vehicle data	Distribution of longitudinal and lateral position at defined time positions
			Vehicle data	Distribution of position in lane

RQ-ID	RQ Level 1	RQ Level 2	Logging requirements	Performance Indicators
RQ-T8		What is the impact of ADF on the driven speed?	Vehicle data	Distribution of velocity
RQ-T9		What are the impacts of ADF on energy efficiency?	Vehicle data	Distribution of fuel consumption and speed
RQ-T10		What is the impact of ADF on the frequency of near-crashes / incidents?	Vehicle data	Frequency of harsh braking
			Vehicle data	Frequency of unintended lane departures
RQ-T11		What is the impact of ADF on the frequency of certain events?	Vehicle data	Frequency of detected driving events
RQ-T12	What is the impact of ADF on the interaction with other road users	What is the impact of ADF on the interaction with other road users in a defined driving scenario?	Vehicle data	Distribution of THW, encroachment time, proportion of stopping distance, post encroachment time, initially attempted post encroachment time
			Vehicle data	Distribution of velocity and acceleration of other road users
			Vehicle data	Distribution of velocity of other road users, Frequency of incidents caused by pedestrians
			Vehicle data	Distribution of THW and TTC
RQ-T13		What are the impacts of ADF on traffic efficiency?	Vehicle data	Distribution of longitudinal velocity, crossing time, waiting time, longitudinal acceleration
RQ-T14			Vehicle data	Frequency of near crashes

RQ-ID	RQ Level 1	RQ Level 2	Logging requirements	Performance Indicators
		What is the impact of ADF on the number of near-crashes / incidents with other road users?	Vehicle data	Frequency of near crashes with VRU
			Vehicle data	Frequency of incidents per country
RQ-T15	What is the impact on the behaviour of other traffic participants?	How does the ADF influence the behaviour of subsequent vehicles?	Vehicle data	Distribution of longitudinal acceleration of other road users
RQ-T16		How does the ADF influence the behaviour of preceding vehicles?	Vehicle data	Frequency of passive cut-in manoeuvres
			Vehicle data	Distribution of THW
RQ-T17		What is the impact of ADF on the number of near-crashes / incidents of other traffic participants?	Vehicle data	Frequency of harsh braking of subsequent vehicles
			Vehicle data	Frequency of very close distance events of subsequent vehicles
RQ-T18		How does the ADF influence the behaviour of other vehicles at intersections?	Vehicle data	

6.2 User & Acceptance Evaluation

RQ-ID	RQ Level 1	RQ Level 2	Logging requirements	Performance Indicators
RQ-U1	What is the impact on user acceptance & awareness?	Are drivers willing to use an ADF?	Proportion of driving time in automated mode, interview/questionnaire for parking function	Proportion of driving time in automated mode. Questionnaire for parking function Annual survey Site-specific questionnaire (ADF use general, specific ADF use)
RQ-U2		How much are drivers willing to pay for the ADF?	Questionnaire	Willingness to pay by questionnaire; for participants and as part of annual survey
RQ-U3		What is the user acceptance of the ADF?	Questionnaire	ADF specific questions to participants (pre, post use at each site); based on pre-existing acceptance questionnaires, tailored to suit automated driving functions Annual survey: non-participants
			Questionnaire	ADF specific questions to participants (pre, post use at each site); based on pre-existing acceptance questionnaires, tailored to suit automated driving functions Annual survey: non-participants
			Questionnaire	ADF specific questions to participants (pre, post use at each site); based on pre-existing acceptance questionnaires, tailored to suit automated driving functions Annual survey: non-participants
			Questionnaire	ADF specific questions to participants (pre, post use at each site); based on pre-existing acceptance questionnaires, tailored to suit automated driving functions

RQ-ID	RQ Level 1	RQ Level 2	Logging requirements	Performance Indicators
				Annual survey: non-participants
			Questionnaire	ADF specific questions to participants (pre, post use at each site); based on pre-existing acceptance questionnaires, tailored to suit automated driving functions Annual survey: non-participants
			Questionnaire	Proportion of driving time in automated mode before and after unexpected use case (requires identification of unexpected use case - based on system status information from the vehicle) ADF specific questionnaire (change in ratings pre- and post- unexpected use case) ADF specific questionnaire on perception of unexpected use cases
RQ-U5		What is the impact of ADF on driver state?	Questionnaires & interviews	ADF specific questions Online physiological measurements possible. Hand position on the steering wheel can be used as a proxy for stress, and could be obtained from video
			Videos and eye-tracking	Subjective data/video coding. Rating of driver video, subjective rating (KSS), eyelid based indicators (PERCLOS)
			Questionnaires	Real-time single scale workload ratings Post-trip workload questionnaires
RQ-U6		What is the impact of ADF use on driver awareness?	Eye tracking, questionnaires	General questionnaire on detection of objects in scene during automated mode and non-automated mode. This would require some 'scripting' of events

RQ-ID	RQ Level 1	RQ Level 2	Logging requirements	Performance Indicators
				Eye-tracking - percent road centre, proportion of time spent looking at specific regions of interest (e.g. mirrors), number of glances and mean/max/SD glance duration would also be useful if eye-tracking available. Some of these could potentially be coded from video - simplistic coding of glance location may be useful as a compromise.
			Eye tracking, questionnaires	Participant questionnaire on perceived risk (does this differ from safety)? Response time to hazards e.g. brake reaction time to lead vehicle or pedestrian-related events
RQ-U4		What are drivers' expectations regarding system features?	Questionnaire	Annual survey - What are drivers’ expectations? ADF specific questions - Did the system meet driver expectations? What could be changed to improve driver understanding of system features - free text responses. Subjective rating of surprise for given driving scenarios?
RQ-U9	What is the user experience?	What is drivers’ secondary task engagement during ADF use?	Driver video (coded for driver secondary task engagement)	Secondary task annotation according to SHRP 2 researchers’ dictionary. Also consider UDRIVE Minimum requirement = code secondary task type. Ideally, include sub-task type e.g. component of a mobile phone task - reach for phone, interact with screen etc. Coding workload would be high, hence could code just a subset of total trip samples
			Video (coded for secondary task engagement, therefore codebook required asap)	Secondary task annotation according to SHRP 2 researchers’ dictionary. Also consider UDRIVE.

RQ-ID	RQ Level 1	RQ Level 2	Logging requirements	Performance Indicators
				Requires coding of secondary task start and stop time, relative to trip duration.
RQ-U10		How do drivers respond when they are required to retake control (in expected and unexpected use cases)?	Questionnaires; vehicle metrics (simulator and test sites)	Time to retake control (simulator and time sites) Success of takeover manoeuvre - steering control immediately after retake of control ADF specific questions relating to takeover experience in specific use cases (perceived safety, perceived quality, confidence)
			Questionnaires; vehicle metrics (simulator and test sites)	Time to retake control (simulator and time sites) Success of takeover manoeuvre - steering control immediately after retake of control ADF specific questions relating to takeover experience in specific use cases (perceived safety, perceived quality, confidence)
RQ-U11		How often and under which circumstances do drivers choose to activate/deactivate the ADF?	Video (coded for system activation deactivation, therefore codebook required asap)	Video coded for environment variables: road type, VRU presence, speed limit, traffic density, weather, lighting conditions. Use UDRIVE codebook for inspiration
RQ-U7		What is the impact of ADF use on motion sickness?	Motion sickness questionnaire	ADF specific questions at pilot sites - Motion sickness Proportion of driving time in automated mode
RQ-U8		What is the impact of motion sickness on ADF use?	Motion sickness questionnaire	ADF specific questions at pilot sites - Motion sickness Proportion of driving time in automated mode

6.3 Impact Evaluation

RQ-ID	RQ Level 1	RQ Level 2	Logging requirements	Performance Indicators
RQ-I1	What is the impact on safety?	What is the impact of ADF on the number of accidents in a certain driving scenario / for certain road users?	Detecting accidents in the simulation (trajectories of the vehicles -> Position and velocity); For setting up the simulation: Driver behaviour of other road users (Pilot --> rel. distance, rel. velocity), Function description (document) or Function behaviour (pilot) In addition: accident data (other data source), traffic data (other data sources)	Frequency of accidents (simulation), Cumulative distribution of per accident type
			Detecting accidents in the simulation (trajectories of the vehicles -> Position and velocity); For setting up the simulation: Driver behaviour of other road users (Pilot --> rel. distance, rel. velocity), Function description (document) or Function behaviour (Pilot) In addition: accident data (other data source), traffic data (other data sources	Frequency of accidents (simulation)
RQ-I2		What is the impact of ADF on accidents with a certain injuries level / damage in a certain driving scenario?	Basically, additional analysis of the detected accidents in RQ-I1; additional data: injury criteria, e.g. injury risk function (other data source)	Frequency of fatal accidents (simulation)

RQ-ID	RQ Level 1	RQ Level 2	Logging requirements	Performance Indicators
			Basically, additional analysis of the detected accidents in RQ-I1; additional data: injury criteria, e.g. injury risk function (other data source)	Frequency of accidents with severe injuries (simulation)
			Basically, additional analysis of the detected accidents in RQ-I1; additional data: injury criteria, e.g. injury risk function (other data source)	Frequency and of accidents with slight injuries (simulation)
			Basically, additional analysis of the detected accidents in RQ-I1; additional data: injury criteria, e.g. injury risk function (other data source)	Frequency of accidents with material damage
			Should build up on the analysis of the accident scenarios required for RQ-I1	
RQ-I3	What is impact of ADF on environmental aspects?	What is the impact on transport network efficiency (throughput) in a certain traffic scenario?		Traffic volume, Traffic flow, Distribution of longitudinal speed over all vehicles, travel time along a specified corridor
				Parking capacity per area
RQ-I4		What is the impact of ADF on energy demand / pollution in a certain traffic scenario?	Fuel consumption of vehicle depending on engine speed and requested torque, velocity, Vehicle data (mass, height, cw,...), speed, acceleration, throttle position, brake pedal position	Average / Amount of fuel consumption per vehicle or over the analysed fleet

RQ-ID	RQ Level 1	RQ Level 2	Logging requirements	Performance Indicators
			Velocity of analyse vehicles (Simulation to be verified by Pilot data), Acceleration, Vehicle data (mass, height, cw,...)	Average / Amount of CO2 per vehicle or over the analysed fleet
			Basically the same as fuel consumption, but different model that provides the CO2 mission	Average / Amount of energy demand (calculated via velocity) per vehicle or over the analysed fleet
RQ-I5	What is the impact of ADF on travel behaviour? (Exposure)	What is the impact of ADF on number of trips made?	Corresponding question	Number of trips per time frame (day, week)
			Corresponding question	Number of parking manoeuvres per time frame (day, week)
RQ-I6		What is the impact of ADF on the frequency of road type usage?	Corresponding question; (Pilot: driven mileage per road type, speed, time)	Proportion of used road types
RQ-I7		What is the impact of ADF on trip duration/distance?	Pilot (time, speed)	Distribution of trip duration and parking duration
			Corresponding question; (Pilot: driven mileage per road type, speed, time)	Distribution of travelled distance
			Pilot: driven speed, detected driving scenario / SV	Distribution of speed, probability density function of driven velocity
RQ-I8		What is the impact of ADF on the frequency of certain driving scenarios (accidents / critical situation / normal driving)?	Pilot (mileage, time, speed, detected driving scenario / situation (aggregated data))	Frequency of accidents
			Pilot (mileage, time, speed, detected driving scenario / situation (aggregated data))	Frequency of driving scenarios

RQ-ID	RQ Level 1	RQ Level 2	Logging requirements	Performance Indicators
RQ-I9		How do the ADF's limitations influence the impact on safety / efficiency?	Function requirements / Information, in which the function did no operate (Pilot); Number of addressed scenarios	

6.4 Socio-Economic Impact Evaluation

RQ-ID	RQ Level 1	RQ Level 2	Logging requirements	Performance Indicators
RQ-S1	What is the socio-economic impacts of ADF?	What is the net welfare gain in a certain societal scenario?	Input from the previous safety & environmental impact assessment Additional information: Cost of functions and vehicle (other sources), Cost of injuries (other sources) / repair, benefit due to additional gain spare time, society data (other sources), Cost due to traffic today (other sources)	Input from the previous safety & environmental impact assessment Additional information: Cost of functions and vehicle (other sources), Cost of injuries (other sources) / repair, benefit due to additional gain spare time, society data (other sources), Cost due to traffic today (other sources)
			Input from the previous safety & environmental impact assessment Additional information: Cost of functions and vehicle (other sources), Cost of injuries (other sources) / repair, benefit due to additional gain spare time, society data (other sources), Cost due to traffic today (other sources)	Input from the previous safety & environmental impact assessment Additional information: Cost of functions and vehicle (other sources), Cost of injuries (other sources) / repair, benefit due to additional gain spare time, society data (other sources), Cost due to traffic today (other sources)
RQ-S2		What is the overall socio-economic impacts for different groups?	Input from the previous safety & environmental impact assessment Additional information: Cost of functions and vehicle (other sources), Cost of injuries (other sources) / repair, benefit due to	Input from the previous safety & environmental impact assessment Additional information: Cost of functions and vehicle (other sources), Cost of injuries (other sources) / repair, benefit due to

RQ-ID	RQ Level 1	RQ Level 2	Logging requirements	Performance Indicators
			additional gain spare time, society data (other sources)	additional gain spare time, society data (other sources)
			Input from the previous safety & environmental impact assessment Additional information: Cost of functions and vehicle (other sources), Cost of injuries (other sources) / repair, benefit due to additional gain spare time, society data (other sources)	Input from the previous safety & environmental impact assessment Additional information: Cost of functions and vehicle (other sources), Cost of injuries (other sources) / repair, benefit due to additional gain spare time, society data (other sources)
			Input from the previous safety & environmental impact assessment Additional information: Cost of functions and vehicle (other sources), Cost of injuries (other sources) / repair, benefit due to additional gain spare time, society data (other sources)	Input from the previous safety & environmental impact assessment Additional information: Cost of functions and vehicle (other sources), Cost of injuries (other sources) / repair, benefit due to additional gain spare time, society data (other sources)

7 Next steps and recommendations

Research questions and hypotheses developed and presented in this document are the starting point for the work that follows in SP3 but also in other SPs. Based on this work, the next steps in SP3 will be:

- WP3.4 will develop the experimental design needed to answer the Research Questions. Due to the variability of research areas a mixture of different experimental procedures will presumably be needed; (reported in D3.2). Datasets from previous projects (e.g. euroFOT) will also be considered to determine whether they can offer useful (baseline) data for the L3Pilot evaluation.
- WP3.5 will collect and describe methods to be used for answering the Research Questions. A range of methods will be needed to effectively answer Research Question covering technical aspects (e.g., change of vehicle behaviour compared to manual driving), user-related concepts (e.g. acceptance), and impact of the ADFs (e.g. on safety, environment, and general societal effects) (reported in D3.3).

Furthermore, the list of Research Questions is directly linked to the work of other SPs in L3Pilot:

- SP4 sets up the vehicles with the different ADFs to be tested in L3Pilot. In case there are changes of the tested ADFs in comparison to the functionalities described in D3.1, an adaptation of Research Questions might become necessary.
- The logging needs derived from the hypotheses are provided to SP5, which will define a common data format to be used throughout the project and set up the tools needed for storing and analysing the data.
- SP6 will run the pilots and will collect the data for answering the Research Questions provided by D3.1.
- The list of Research Questions from D3.1 will be the basis for the analysis to be done in SP7 at the end of the project. The aim of SP7 will be to test the hypotheses derived from the Research Questions presented in D3.1.

Research Questions provide the foundation for L3Pilot and have an impact on the entire project in the coming years. Due to the complexity of the whole process of implementing the ADFs, running the pilots and putting the needed methods into place, some Research Questions will be adapted later on in the project. Any changes to the list of Research Question will be documented in D.3.4.

References

Abou-Zeid, M., Witter, R., Bierlaire, M., Kaufmann, V. & Ben-Akiva, M. (2012). Happiness and travel mode switching: Findings from a Swiss public transportation experiment. Transport Policy 19, 93–104.

Bagheri, N., & Jamieson, G.A. (2004). *Considering subjective trust and monitoring behavior in assessing automation-induced "complacency ."* In D. A. Vicenzi, M. Mouloua, & O. A. Hancock (Eds.), Human performance, situation awareness, and automation: Current research and trends (pp. 54-59). Mahwah, NJ: Erlbaum.

Bartels, A., Eberle, U. & Knapp, A (2015), *System Classifications and Glossary*, AdaptIVE.

R.Behringer, "The DARPA Grand Challenge - Autonomous Road Vehicles in the Desert," in Proc. Intelligent and Autonomous Vehicles (IAV), Lisbon, 2004.

Beller, J., Heesen, M. and Vollrath, M. (2013). Improving the Driver-Automation Interaction: An Approach Using Automation Uncertainty. Human Factors, 55(6), 1130-1141.

Billings, C. E. (1997). *Aviation Automation: The Search for a Human-Centered Approach*. Mahwah, NJ: Chicago, IL: University of Chicago Press.

Biondi, F., Strayer, D.L., and Drews, F. (2014). Autonomous vehicle: are you aware of them? In: *Proceedings Human Factors and Ergonomics Society conference*. Lisbon, Portugal.

Brignolo, R., Rämä, P., Cosenza, S., Innamaa, S., Gasser, T. Rappold, N., Le, L., Bergman, J., Schindhelm, R., Benz, T., Brouwer, R., Jonkers, E. & Möller, M. (2011). Drive C2X FOT research questions and hypotheses and experimental design. Deliverable D42.1 of DriveC2X project. Version 1.4.

Button, K., Stough, R., Bragg, M. & Taylor, S. (2006). Telecommunications, Transportation and Location. Massachusetts. Edward Elgar Publishing, Inc. Carlson & Marchi 2014

https://www.theverge.com/2017/10/11/16458850/self-driving-car-california-dmv-regulations

https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/testing

Carsten, O. and Nilsson, L. (2001). Safety assessment of driver assistance systems. *European Journal of Transport and Infrastructure Research*, 1(3), 225-243.

CARTRE (2017), "User Awareness, Users and Societal Acceptance and Ethics, Driver Training: Position Paper", *CARTRE Thematic Interest Group*, CARTRE, unpublished.

Chaparro, A., Groff, L., Tabor, K., Sifrit, K., and Gugerty, L. J. (1999, September). Maintaining situational awareness: The role of visual attention. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 43, No. 23, pp. 1343-1347). Sage CA: Los Angeles, CA: SAGE Publications.

Collet, C., Petit, C., Champely, S. and Dittmar, A. (2003). Assessing workload through physiological measurements in bus drivers using an automated system during docking. *Human Factors*, 45, 539-548.

Collins Dictionary (n.d.), *Pilot Study*, Available from: https://www.collinsdictionary.com/dictionary/english/pilot-study.

Dambock, D., Farid, M., Tonert, L. and Bengler, K. (2012). Ubernahmezeiten beim hochautomatisierten Fahren. Paper presented at 5. Tagung Fahrerassistenz, Munchen.

Davis, J., Animashaun, A., Schoenherr, E., and McDowell, K. (2008). Evaluation of semi-autonomous convoy driving. *Journal of Field Robotics*, 25, 880-897.

Department for Transport (2016).

De Waard, D. (1996). *The measurement of drivers' mental workload*. PhD thesis. University of Groningen, Groningen, Netherlands.

De Winter, J. C. F., Happee, R., Martens, M. H. and Stanton, N. A. (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*, 27 Part B, 196-217.

Eisses, S. (2011). ITS Action Plan. D4-Final report. Action 3.4-Safety and comfort of the Vulnerable Road User.

Endsley, M. R., Sollenberger, R., and Stein, E. (2000). Situation awareness: A comparison of measures. *Proceedings of the human performance, situation awareness and automation: user-centered design for the new millennium*, 15-19.

Endsley, M.R. (1988) Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society 32nd Annual Meeting, Human Factors Society*, Santa Monica, CA, 97-101.

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32-64.

European Commission (2014). Guide to Cost-Benefit Analysis of Investment Projects http://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/cba_guide.pdf

Flemisch, F., Heesen, M., Hesse, T., Kelsch, J., Schieben, A. and Beller, J. (2012). Towards a dynamic balance between humans and automation: Authority, ability, responsibility and control in cooperative control situations. *Cognition, Technology and Work*, 14, 3-18.

FOT-Net (2017). FESTA Handbook, version 7. http://fot-net.eu/Documents/festa-handbook-version-7/

FOT-Net Data (2016), *FESTA Handbook: Version 6*.

Gärling, T. & Axhausen, K. (2003). Introduction: Habitual travel choice. Transportation 30 (1), 1–11.

Gelau, C.; Vierkötter, M.; Rakic, M., Bärgman, J.; Kotiranta, R.; Victor, T.; Regan, M.; Saad F.; Engel, R.; Page, Y.; Rognin, L.; Brouwer, R.; Feenstra, P.; Hogema, J.; Malone, K.; Minett, C.; Schrijver, J.; Barnard, Y.; Jamson, H.; Jamson, S.; Hammarström, U.; Hjälmdahl, M.; Kircher, A.; Kircher, K.; Tapani. A.; Luoma, J.; Peltola, H. & Rämä, P. 2008. A comprehensive framework of performance indicators and their interaction. Deliverable 2.1 of FESTA Support Action. 7th Framework programme.

Gold, C., Dambock, D., Lorenz, L., and Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; 57, 1938-1942.

Gouy, M., Wiedemann, K., Stevens, A., Brunett, G., & Reed, N. (2014). Driving next to automated vehicle platoons: How do short time headways influence non-platoon drivers' longitudinal control? Transportation Research Part F: Traffic Psychology and Behaviour, 27(PB), 264–273. https://doi.org/10.1016/j.trf.2014.03.003

Green, P. (1999). *Visual and task demands of driver information systems* (No. UMTRI-98-16). Ann Arbor: University of Michigan Transportation Research Institute.

Gudmundsson, H. (2005). Mobility as a Policy Concept. In: Thomsen, T., Nielsen, L. & Gudmundsson, H. (ed.) Social Perspectives on Mobility. Ashgate Publishing Limited, Hampshire, 107–126.

Hakonen, M. (2010). Validation of a mobility impact assessment method. Master of science thesis, Tampere University of Technology.

Hanson, S. (1995). Getting there: urban transportation in context. Geography of urban transportation.

HEATCO (2006). Developing harmonised European approaches for transport costing and project assessment. Available from http://heatco.ier.uni-stuttgart.de

Hergeth, S., Lorenz, L., & Krems, J. F. (2017). Prior familiarization with takeover requests affects drivers' takeover performance and automation trust. *Human factors*, 59(3), 457-470.

Hoedemaeker, M. (2000). Driving with intelligent vehicles. In IEEE Intelligent Transportation Systems Conference (pp. 506–509). Dearborn.

Innamaa, S., Smith S., Barnard, Y., Rainville, L., Rakoff, H., Horiguchi, R. & Gellerman, H. (2017). Trilateral impact assessment framework for automation in road transportation. Draft 4 January 2017. ART WG.

Innamaa, S., Axelson-Fisk, M., Borgarello, L., Brignolo, R., Guidotti, L., Martin Perez, O., Morris, A., Paglé, K., Rämä, A., Wallgren, P. & Will, D. (2013). Impacts on Mobility – Results and Implications. Large Scale Collaborative Project, 7th Framework Programme, INFSO-ICT 224067. European Commission 2013 No: Deliverable D4.4.3.

ITF (2017), Quantifying the Socio-economic Benefits of Transport, ITF Roundtable Reports, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789282108093-en

Jamson, A. H., Merat, N., Carsten, O. M. J. and Lai, F. C. H. (2013). Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation Research Part C: Emerging Technologies*, 116-125.

Jamson, A. H., Merat, N., Carsten, O. M. J., & Lai, F. C. H. (2013). Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. Transportation Research Part C: Emerging Technologies, 30, 116–125. https://doi.org/10.1016/j.trc.2013.02.008

Jin, S. (2008). *The Effect of Driver Cognitive Abilities and Distraction on Situation Awareness and Performance under Hazard Conditions*. Master's Thesis. North Carolina State University, USA.

Kellerman, A. (2016). Daily spatial mobilities: Physical and virtual. Routledge.

Kircher, K., Larsson, A. and Hultgren, J. A. (2014). Tactical Driving Behavior With Different Levels of Automation. *IEEE Transactions on Intelligent Transportation Systems*, 15(1), 158-167.

Klöckner, C. & Friedrichsmeier, T. (2011). A multi-level approach to travel mode choice – How person characteristics and situation specific aspects determine car use in a student sample. Transportation Research Part F: Traffic Psychology and Behaviour 14 (4), 261–277.

Korbel, P., Skulimowski, P., Wasilewski, P. & Wawrzyniak, P. (2013). Mobile applications aiding the visually impaired in travelling with public transport. Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on. IEEE 2013, 825–828.

Kulmala, R. & Rämä, P. (2010). Mobility Data Analysis Plan. Large Scale Collaborative Project, 7th Framework Programme, INFSO-ICT 224067. European Commission 2013. Deliverable T4.4.1.

Kulmala, R. (2010). Ex-ante assessment of the safety effects of intelligent transport systems. Accident Analysis and Prevention, 42, 1359–1369.

Larsson, A. F. (2013). *Automation and the nature of driving. The effect of adaptive cruise control on drivers' tactical driving decisions* [Dissertation]. Lund, Lund University.

Larsson, A. F. L., Kircher, K., & Hultgren, J. A. (2014). Learning from experience: Familiarity with ACC and responding to a cut-in situation in automated driving. Transportation Research Part F: Traffic Psychology and Behaviour, 27(PB), 229–237. https://doi.org/10.1016/j.trf.2014.05.008

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

Lee, J. D., McGehee, D. V., Brown, T. L., and Reyes, M. L. (2002). *Driver distraction, warning algorithm parameters, and driver response to imminent rear-end collisions in a high-fidelity driving simulator* (No. DOT-HS-809-448). Chicago

Llaneras, R. E., Salinger, J., and Green, C. A. (2013). Human factors issues associated with limited ability autonomous driving systems: Drivers' allocation of visual attention to the forward roadway. In *Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment*, Training and Vehicle Design (pp. 92-98).

Lorenz, L, Kerschbaum P. and Schumann, J. (2014). Designing take over scenarios for automated driving: How does augmented reality support the driver to get back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting September 2014*, 58: 1681-1685, doi:10.1177/1541931214581351

Louw, T., Madigan, R., Carsten, O., & Merat, N. (2016). Were they in the loop during automated driving? Links between visual attention and crash potential. *Injury prevention,* 23, 281-286.

Louw, T., & Merat, N. (2017). Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transportation Research Part C: Emerging Technologies*, 76, 35-50.

Louw, T., Markkula, G., Boer, E., Madigan, R., Carsten, O., & Merat, N. (2017). Coming back into the loop: Drivers' perceptual-motor performance in critical events after automated driving. *Accident Analysis & Prevention*, 108, 9-18.

Mäkinen, T., Schulze, M., Krajcewicz, D., Gaugel, T. & Koskinen, S. (2011), *DRIVE C2X methodology framework*, DRIVE C2X.

Malone, K., Silla, A., Johansson, C., and Bell, D. (2017). Safety, mobility and comfort assessment methodologies of intelligent transport systems for vulnerable road users. European Transport Research Review, 9:21.

Malone, K., Rech, J., Hogema, J., Innamaa, S., Hausberger, S., Dippold, M., van Noort, M., de Feijter, E. Rämä, P., Aittoniemi, E., Benz, T., Burckert, A., Enigk, H., Giosan, I. Gotschol, C., Gustafsson, D., Heinig, I., Katsaros, K., Neef, D., Ojeda, L., Schindhelm, R., Sütterlin, C. & Visintainer, F. (2014). Impact Assessment and User Perception of Cooperative Systems. Large-scale intergrating project, 7th Framework programme. Objective ICT-2009.6.2: ICT for Mobility of the Future. Deliverable D11.4.

Malta, L., Ljung Aust, M., Faber, F., Metz, B., Saint Pierre, G., Benmimoun, M. & Schäfer, R. (2012). Final results: Impacts on traffic safety. Deliverable 6.4 of euroFOT project. 7th Framework programme.

Manca, L., De Winter, J., and Happee, R. (2015). Visual displays for automated driving: A survey. In Proceedings of the *7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (AutoUI 2015) (pp. 1-5). New York, NY: ACM.

Manser, M.P., Creaser, J., and Boyle, L.N. (2012). Behavioral Adaptation. Methodological and Measurement Issues. In: Rudin-Brown, C. & Jamson, S.L. (Eds.): Behavioural Adaptation to Road Safety. Theory, Evidence and Action (pp. 340-357). Boca Raton: Taylor & Francis Group.

Martens, M.H. & Jenssen, G.D. (2012). Behavioral Adaptation and Acceptance. In: Eskandarian, A. (Ed.), *Handbook of Intelligent Vehicles* (pp. 117-138). Washington, DC: Springer.

Merat, N. and Jamson, A. H. (2008). How do drivers behave in a highly automated car? In *Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment*, Training and Vehicle Design (pp. 514-521). Big Sky, MT.

Merat, N. and Jamson, A. H. (2009). *Is drivers' situation awareness influenced by a fully automated driving scenario?* In D. de Waard, J. Godthelp, F. L. Kooi, and K. A. Brookhuis (Eds.), Human factors, security and safety. Maastricht, the Netherlands: Shaker Publishing.

Merat, N. and Lee, J. D. (2012). Preface to the special section on human factors and automation in vehicles designing highly automated vehicles with the driver in mind. *Human Factors*, 54, 681-686.

Merat, N., Jamson, A. H., Lai, F. C. H., Daly, M., and Carsten, O. M. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*, 27, 274-282.

Methorst, R. (2003). Assessing Pedestrians' Needs. The European COST 358 PQN project.

Metz, D. (2000). Mobility of older people and their quality of life. Transport Policy 7, 149–152.

Milakis, D., van Arem, B. & van Wee, B. (2017). Policy and society related implications of automated driving: A review of literate and directions for future research. Journal of Intelligent Transportation systems, 21 (4), 324–348.

Mononen, P., Franzen, S., Pagle, K., Morris, A., Innamaa, S., Karlsson, M., Touliou, K., Montanari, R., Fruttaldo, S. (2012). Final Report. TeleFOT deliverable D1.15. 156 p.

Moray, N. (1990). Designing for transportation safety in the light of perception, attention, and mental models. *Ergonomics*, 33(10), 1201-1213.

Murtagh, N., Gatersleben, B. & Uzzell, D. (2012). Self-identity threat and resistance to change: Evidence from regular travel behaviour. Journal of Environmental Psychology 32, 318–326.

Naujoks, F., Mai, C. and Neukum, A. (2014). The effect of urgency of take-over requests during highly automated driving under distraction conditions. In T. Ahram, W. Karowski and T. Marek (Eds.), *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE 2014* (pp. 2099-2106). Krakau: AHFE Conference.

Neubauer, C., Matthews, G., and Saxby, D. (2012). The effects of cell phone use and automation on driver performance and subjective state in simulated driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56(1), 1987-1991.

Noort, M. van, Malone, K., Silla, A., Leden, L., Rämä, P., Innamaa, S., Johansson, C., Bell, D., Giannelos, I., Mans, D., van Schijndel, M. and Morris, A. (2009). Assessment Methodology D2.2 of VRUITS.

Norros, L. (2004). Acting under uncertainty: The core-task analysis in ecological study of work. VTT Technical Research Centre of Finland.

OECD (2017), ITF Transport Outlook 2017, OECD Publishing, Paris.

Parasuraman, R., and Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381-410.

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

Paulssen, M., Temme, D., Vij, A. & Walker, J. (2014). Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice. Transportation 41 (4), 873–888.

Payre, W., Cestac, J., and Delhomme, P. (2016). Fully Automated Driving Impact of Trust and Practice on Manual Control Recovery. *Human Factors*, 58(2), 229-241.

Petermeijer, S., De Winter, J. C., and Bengler, K. (2015). Vibrotactile displays: A survey with a view on highly automated driving. *IEEE Transactions on Intelligent Transportation Systems*, 1-11.

Radlmayr, J., Gold, C., Lorenz, L., Farid, M. and Benlger, K. (2014). How Traffic Situations and Non-Driving Related Tasks Affect the Take-Over Quality in Highly Automated Driving. In *Proceedings of the Human Factors and Ergonomics Society (HFES) 2014. Annual Meeting* (Vol. 58, pp. 2063-2067).

Rämä, P. & Koskinen, H. (2017). Exploring Driver and Operator Behaviour Models in the Context of Automated Driving – Identification of Issues from Human Actor Perspective. Draft 27.8.2017.

Rauch, N. (2009). *Ein verhaltensbasiertes messmodell zur erfassung von situationsbewusstsein im fahrkontext* (doctoral dissertation). Psychologisches Institut der Universitat Wurzburg, Wurzburg, Germany.

Rodarius, C., Duflis, J., Fahrenkrog, F., et al., (2015). Test and evaluation plan. AdaptIVe Deliverable 7.1.

Society of Automotive Engineers [SAE], (2016). *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems (Standard No. J3016A)*. SAE International. Retrieved from: <http://standards.sae.org/j3016_201401/>.

Sager, T. (2006). Freedom as Mobility: Implications of the Distinction between Actual and Potential Travelling. Mobilities 1 (3), 465–488.

Sarter, N.B., Mumaw, R.J., and Wickens, C.D. (2007). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data. *Human Factors*, 49(3), 347-357.

Schulze M., Mäkinen T., Kessel T., Metzner S., Stoyanov, H. (2014). DRIVE C2X Final Report (DRIVing implementation and Evaluation of C2X technology in Europe). Deliverable 11.6. EC 7th Framework Program.

Shaheen, S., Cohen, A., Zohdy, I. & Kock, B. (2016). Smartphone Applications to Influence Travel Choices: Practices and Policies. U.S. Department of Transportation, Federal Highway Administration, Washington, DC.

Slater, K. (1985) Human comfort. Charles C. Thomas Publisher, Springfield.

Spinney, J., Scott, D. & Newbold, K. (2009). Transport mobility benefits and quality of life: A time-use perspevtive of elderly Canadians. Transport policy 16, 1–11.

Stanton, N. A., and Young, M. S. (2005). Driver behaviour with adaptive cruise control. *Ergonomics*, 48(10), 1294-1313.

Stevens, A. (2004). The application and limitations of cost-benefit assessment (CBA) for intelligent transport systems, in Economic impacts of Intelligent transport systems: Innovations and Case studies in Transportation Economics, Volume 8, 91-111. Elsevier.

Stevens, A., Ball S., Hopkin J. (2016). Evaluating cooperative ITS. Evaluation of Intelligent Road Transport Systems: Methods and Results. IET 2016.

Susilo, Y. O. & Cats, O. (2014). Exploring key determinants of travel satisfaction for multi-modal trips by different traveler groups. Transportation Research Part A: Policy and Practice 67, 366–380.

Thaler, R. & Sunstein C. (2008). Nudge: Improving Decisions About Health, Wealth, and Happiness. Yale University Press, New Haven & London.

van den Beukel, A. P., and van der Voort, M. C. (2013, October). The influence of time-criticality on situation awareness when retrieving human control after automated driving. In *Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference* (pp. 2000-2005). IEEE.

Van Wee, B., Holwerda, H. & Van Baren, R. (2002). Preferences for modes, residental location and travel behaviour: The relevance for land-use impacts on mobility. European Journal of Transport and Infrastructure Research 2 (3/4), 305–316.

Welsh, R., Morris, A., Reed, S., Wallgren, P., Innamaa, S., Rämä, A., Perez, O., Birell, S., Will, D., Brignolo, R. & Guidotti, L. (2008). Impacts on Safety – Results and Implications. D4.4.3 of TeleFOT project.

Whitlow, S. D., Dorneich, M. C., Funk, H. B. and Miller C. A. (2002). Providing appropriate situation awareness within a mixed-initiative control system. In 2002 IEEE International Conference on Systems, Man and Cybernetics, volume 5, October 2002.

Young, M. S., & Stanton, N. A. (2007). Back to the future: Brake reaction times for manual and automated vehicles. Ergonomics, 50(1), 46–58. https://doi.org/10.1080/00140130600980789

Zeeb, K., Buchner, A., and Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accident Analysis & Prevention*, 88, 212-221.

List of abbreviations and acronyms

Abbreviation	Meaning
AD	Automated Driving
ADAS	Advanced Driver Assistance Systems
ADF	Automated Driving Function
ADS	Automated Driving System
AV	Automated Vehicle
CAN	Controller Area Network
CBA	Cost Benefit Analysis
DAS	Data Acquisition System
FOT	Field Operational Test
GPS	Global Positioning System
HMI	Human-Machine Interaction
NDD	Naturalistic Driving Data
NDS	Naturalistic Driving Study
ODD	Operational Design Domain
PI	Performance Indicator
RQ	Research Question
SAE	Society of Automotive Engineers
SAE L3	SAE Level 3
SCE	Safety Critical Event
SP	Sub Project
TOR	Take Over Request
WoZ	Wizard of Oz (Method)
WP	Work Package