Deliverable **D7.2** /

L3/L4 long-term study about user experiences

Version: 2.0 Final

Dissemination level: PU

Lead contractor: WIVW GmbH

Due date: 31.10.2021

Version date: 30.09.2021

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

Document information

Authors

Barbara Metz – WIVW GmbH
Johanna Wörle – WIVW GmbH
Alexander Zerbe – BASt
Roland Schindhelm – BASt
Alexander Frey – BASt
Frank Bonarens – Opel
Tyron Louw – University of Leeds
Yee Mun Lee – University of Leeds
Ruth Madigan – University of Leeds
Natasha Merat – University of Leeds
Linda Pipkorn – Chalmers
Thomas Streubel – Chalmers
Emma Tivesten – VCC

Coordinator

Aria Etemad
Volkswagen Group Innovation
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

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

Abbreviation	Meaning
ACC	Adaptive Cruise Control
AD	Automated Driving
ADAS	Advanced Driver Assistance Systems
ADF	Automated Driving Function
BA	Behavioural Adaptation
DMS	Driver Monitoring System
GF	Off-road glange frequency
FOT	Field Operational Test
HMI	Human-Machine Interface
HYP	Hypothesis
ISA	Intelligent Speed Adapter
KSS	Karolinska Sleepiness Scale
MRM	Minimal Risk Manoeuvre
NDRA	Non-Driving Related Activity
ODD	Operational Design Domain
PI	Performance Indicator
PERCLOS	Percentage eyelid closure
PRC	Percent road centre
RQ	Research Question
SAE	Society of Automotive Engineers
L2	SAE Level 2
L3	SAE Level 3
L4	SAE Level 4
SP	Sub Project
TOC	Take-over-controllability
TOR	Take-over request

Executive Summary

The deliverable “D7.2 L3/L4 long-term study about user experiences” describes work within L3Pilot that investigated the change of user experience with long-term usage of an ADF. In the literature, this topic is summarised under the broader concept of behavioural adaptation. During the project it turned out that there are unanswered questions regarding behavioural adaptation in addition to the question of system perception that is built on repeated exposure. Due to that, the scope of this deliverable has been broadened to studies conducted in L3Pilot that deal with behavioural adaptation independent of the considered time span and closely related user topics.

The deliverable starts with an overview of reported research and a short look at the literature. Then, eight chapters summarise eight supplementary studies which addressed user related topics by using methods that go beyond the on-road pilot testing in L3Pilot. Three studies investigated the change of acceptance and usage with repeated usage of a motorway chauffeur. One study took place in a driving simulator, one used a Wizard of Oz vehicle and one addressed the topic as part of the on-road pilot testing. Furthermore, there is work focusing on short-tem behavioural adaptation, i.e. an immediate impact of driving with an ADF on manual driving behaviour directly after or during the transition of control. Lastly, drivers’ preferences with regard to potential side tasks while driving with an automated driving system were explored in an online-survey.

Overall, it was shown that when using an automated driving functions on multiple drives, drivers get used to the system, they gain trust and evaluate the function positively. The studies on short-term behavioural adaptation indicate that further research is needed to better understand whether and how driving with an automated driving function impacts manual driving behaviour immediately after transferring control back to the driver.

For future work it is recommended to further investigate the topic of behavioural adaptation in the long- but also the short-term perspective. This will help to design automated driving systems in way that drivers feel like using the function and that usage is safe right from the beginning and stays safe after getting used to the system.

1 Introduction

1.1 Motivation for 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 it is at a stage that justifies automated driving tests in large-scale pilots. The issues of automation are not solved simply by integrating more and better technology. Above all, this topic needs a focus on user behaviour with automated driving systems. User acceptance is a key factor in the uptake of AD, as is an understanding of the technological, infrastructural, social, and legal aspects that first need to be discussed and resolved on a broad level.

L3Pilot is taking important steps towards the introduction of automated cars in daily traffic. The project undertook large-scale testing and piloting of AD with developed SAE Level 3 (L3) functions (Figure 1.1) in on-road tests across Europe. Level 4 (L4) functions were also assessed in some cases. It should be noted that an important distinction between Level 2 (L2) and L3 systems is the shift in monitoring responsibility from the human to the AD system (SAE, 2021). With an L2 function, the responsibility is on the driver to monitor the driving task and driving environment constantly. With an L3 function, the driving task is performed by the vehicle, but the driver has to remain ready to take-over control if necessary. This difference means that there is a considerable change in the technical capabilities of an L3 automated driving function (ADF) compared to L2. The overall objective of the L3Pilot project was to test and study the viability of automated driving as a safe and efficient means of transportation and to explore and promote new service concepts to provide inclusive mobility.

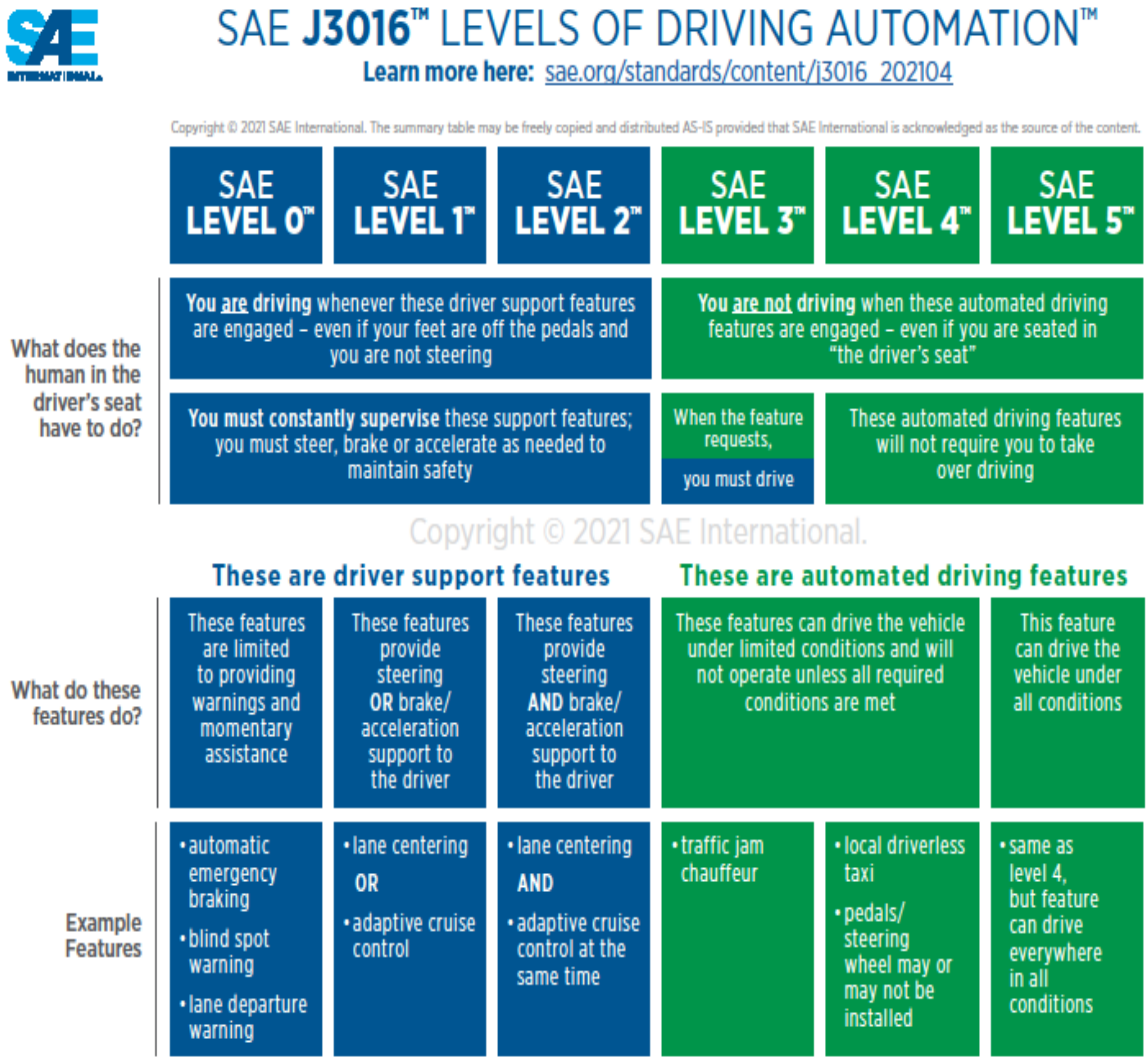

Figure 1.1: SAE Levels of Driving Automation J3016 (Copyright 2021 SAE International).

1.2 Approach and Scope

The L3Pilot project comprised the large-scale piloting of ADFs, primarily L3 functions, with additional assessment of some L4 functions. The key in testing is to ensure that the functionality of the systems used is exposed to variable conditions and that performance is consistent, reliable, and predictable, as this will facilitate a positive experience for the user. A good experience of using AD can accelerate acceptance and adoption of the technology and improve the business case to deploy AD.

The work in L3Pilot was structured into different subprojects that dealt with different aspects. The project's structure followed the FESTA V-process methodology (FOT-Net, 2018) of setting up and implementing tests, adapting the methodology to suit L3Pilot needs as follows:

(i) Prepare: SP3, SP4, SP5

(ii) Drive: SP6

(iii) Evaluate: SP7

As part of the preparation phase, functions and use cases were determined and research questions (RQs) and hypotheses (HYPs) were formulated and reported in D3.1. Here, it was stated that the piloting would mainly focus on RQs and HYPs in four impact areas: (i) safety, (ii) mobility, (iii) efficiency, and (iv) environment. Additional evaluation areas were addressed as well and included issues such as user and acceptance evaluation.

For the area of user-related aspects, it turned out that not all RQs could be studied in on-road pilot tests, mainly due to safety constraints and practical reasons. This especially concerned changes in perception manifesting with repeated usage and growing familiarity with an L3 function of high relevance. Compared to studies with a single experience of L3 functions, perception built on repeated experience provides more reliable insights into the likely acceptance and usage of L3 functions after market introduction. Due to the prototype nature of the functions tested in the on-road study, the assessment of repeated usage of L3 functions by ordinary drivers in real traffic presented a challenge in L3Pilot.

1.3 Introduction to Deliverable 7.2

This deliverable "D7.2 L3/L4 long-term study about user experiences" was planned to describe work within L3Pilot that investigated changes in user experiences after long-term usage of an ADF. In the literature, this topic is summarised under the broader concept of behavioural adapation. The project showed that there is currently no clear definition of "long-term" usage in comparison to "short-term" usage. Furthermore, it suggested that there are unanswered questions with regard to behavioural adaptation in addition to the impact of long-term usage on system perception. Because of this, the scope of this deliverable was extended to include studies that would deal with behavioural adaptation independent of the time span of usage and closely related user topics.

This deliverable focuses on supplementary studies conducted within L3Pilot in addition to the on-road studies reported in other delivarables. All supplementary studies dealt with user-related topics that were difficult to address in the on-road tests of the project. Due to safety requirements and legal issues, the prototype nature of ADFs tested in the on-road tests of L3Pilot made it difficult to use ordinary drivers as test participants. Therefore, supplementary studies were planned in addition to the on-road tests, to be able to study system usage and other relevant user topics with ordinary, non-professional drivers in a safe environment. One major topic of the supplementary studies was behavioural and attitudinal changes that occur with the actual usage of ADF. This type of research provides insights into how the usage of ADFs changes over time and how the users' perception of the system changes over time. Drivers using an ADF for the first time (as in most on-road studies in L3Pilot) focus their attention on the system and concentrate on system behaviour. With growing experience, drivers become accustomed to the system and their usage and experience of the system changes. This change over time is especially important to consider when future ADF usage is being estimated. Furthermore, it was assessed how ADF use impacts manual driving behaviour immediately after a transition of control.

2 Overview of supplementary studies – objectives & research questions

2.1 Objectives

Due to the prototype nature of the ADFs tested in the pilot tests, not all research questions on user-related aspects could be addressed in the on-road tests. Therefore, several supplementary studies were conducted in L3Pilot which focused primarily on specific additional user-related research questions. Table 2.1 gives an overview of the supplementary studies reported in this deliverable.

- Three supplementary studies investigated – amongst other concepts – the change of ADF usage and acceptance with repeated experience of the system. This change is also referred to as behavioural adaptation (BA) with a more long-term perspective.
- Two studies in the driving simulator and several on-road studies with a Wizard of Oz vehicle investigated short-term changes of manual driving behaviour after driving with an ADF and the impact of ADF design. These topics are also referred to as short-term BA.
- The impact of ADF use on drivers' fatigue was assessed in a wizard of Oz study taking place on a test track.
- In an online survey, the drivers' expectation regarding secondary task interaction while using an ADF was investigated.

Table 2.1: Overview of supplementary studies.

Study Nr	Study title	Main Topic	Adressed SAE level	Approach	Further description	Related chapter
1	Wizard of Oz study on long-term behavioural adaption	Behavioural adaptation – long-term	L3	Wizard of Oz	Case study, Repeated usage (3 times)	Chapter 4
2	Driving simulator study on long-term behavioural adaption	Behavioural adaptation – long-term	L3 & L4	Simulator	Repeated usage (6 times)	Chapter 5
3	On-road study on on long-term behavioural adaption	Behavioural adaptation – long-term	L4	On-road study	Repeated usage (3 times)	Chapter 6
4	Driving simulator study on short-term behavioural adaption	Behavioural adaptation – short-term	L3	Simulator	Impact of ADF on manual driving	Chapter 9
5	Driving simulator study on ambient peripheral light display	Behavioural adaptation – short-term	L3	Simulator	Impact of HMI on take-over response and acceptance	Chapter 10
6	Wizard of Oz studies on take-over performance	Behavioural adaptation – short-term	L2 & L3	Wizard of Oz, test track & Wizard of Oz, public road	Variation of take-over situation Take-over response process N=4 studies	Chapter 7
7	Driver impairment study	Fatigue & Alcohol (BAC ≈ 0.1%)	L0, L2, & L3.	Wizard of Oz, test track	Sleepiness and visual attention	Chapter 8
8	Online study on User Acceptance	NDRA engagement		Survey	Secondary task interruption by TOR	Chapter 0

Based on the described objectives, specific RQs were developed for the supplementary studies that start with the common L3Pilot RQs (see Hibberd et al. 2018), but go beyond them. For instance, the RQs on long-term effects cover the same research areas as the RQs on user-related aspects defined in D3.1 (see Hibberd et al. 2018). However, for the long-term studies the focus of the RQs is not on comparing ADF with manual driving but rather on assessing user topics regarding ADF over time.

2.2 Research questions adressed

Table 2.2 lists the RQs defined in Hibberd et al. (2018). The third column contains adapted hypotheses which are addressed in at least one of the supplementary studies. The ID of the RQ provides a direct link to D3.1 (Hibberd et al. 2018) and other deliverables on methods. Besides adapted versions of the common RQs (e.g. by focusing on changes occurring with repeated usage of an ADF), there are also hypotheses that are addressed in the supplementary studies and that go beyond the common RQs. Table 2.2 lists the supplementary studies in which every adapted RQ is addressed.

Table 2.2: Research questions that are addressed in the supplementary studies. WoZ = Wizard of Oz study, DS = driving simulator study.

ID	Common RQ	Specific hypotheses	Study Nr & short title							
			1	2	3	4	5	6	7	8
			WoZ long-term	DS long-term	On-road long-term	DS short-term	DS ambient light	WoZ take-over	Impairmens	Survey
RQ-U1	Are drivers willing to use an ADF?	Willingness to use increases with increasing experience with function.	X	X	X					
RQ-U3	What is the perceived safety of the ADF?	Perceived safety increases with increasing experience with function.		X	X					
	What is the perceived comfort of the ADF?	Perceived comfort increases with increasing experience with function.		X	X					
	What is the perceived reliability of the ADF?	Perceived reliability increases with increasing experience with function.		X	X					
	What is the perceived usefulness of the ADF?	Perceived usefulness increases with increasing experience with function.		X	X				X	X
	What is the perceived trust of the ADF?	Trust increases with increasing experience with function.	X	X	X					
RQ-U4	What are drivers' expectations regarding system features?	With increasing experience, understanding of the system increases.		X						

ID	Common RQ	Specific hypotheses	Study Nr & short title							
			1	2	3	4	5	6	7	8
			WoZ long-term	DS long-term	On-road long-term	DS short-term	DS ambient light	WoZ take-over	Impairmens	Survey
RQ-U5	What is the effect of ADF use on drivers' level of stress?	Over AD usage time, drivers experience less stress.		X						
	What is drivers' level of fatigue while using the ADF?	After a familiarisation period, drivers will become drowsy more rapidly.		X						
		ADF level impacts the development of fatigue.		X				X	X	
	What is drivers' level of workload while using the ADF?	Over AD usage time, drivers experience less workload.		X	X					
RQ-U6	What is the effect of ADF use on driver attention to the road/other road users?	With increasing experience, attention to other road users decreases.	x	X	X				X.	
RQ-U9	What is the frequency and duration of drivers' secondary task engagement during ADF use?	Secondary task interaction increases with increasing experience with function.	X	X						
		Which secondary tasks are preferred by the drivers?		X						X
		HMI-design impacts NDRA engagement.					X			

ID	Common RQ	Specific hypotheses	Study Nr & short title							
			1	2	3	4	5	6	7	8
			WoZ long-term	DS long-term	On-road long-term	DS short-term	DS ambient light	WoZ take-over	Impairmens	Survey
RQ-U10	How do drivers respond when they are required to retake control in take-over situations?	Take-over performance increases with increasing experience with function.	X	X						
		Duration of driving with ADF impacts take-over performance.						X		
		HMI-design impacts take-over performance.					X			
RQ-U11	How often and under which circumstances do drivers choose to activate / deactivate the ADF?	Pattern of system activation will become more dependent on driving scenario with increasing experience with function.	X	X						
RQ-UE1	After ADF use, manual driving behaviour changes.	Behaviour of the ADF impacts manual driving behaviour				X				
		Durarion of ADF use impacts manual driving behaviour						X		
RQ-UE2	ADF level impacts evaluation of the ADF.	Drivers prefer higher level ADFs.		X						

ID	Common RQ	Specific hypotheses	Study Nr & short title							
			1	2	3	4	5	6	7	8
			WoZ long-term	DS long-term	On-road long-term	DS short-term	DS ambient light	WoZ take-over	Impairmens	Survey
RQ-UE3	ADF use by fatigued drivers.	Fatigued drivers accept and use an ADF more than non fatigued drivers.		X						
		Fatigued drivers use driving with L4-ADF to sleep.		X						
RQ-U5E1	Effect of alcohol intoxication on driver sleepiness as a function of drive time at different levels of automation								X	
RQ-U6E1	Effect of alcohol intoxication on driver attention at different levels of automation								X	

2.3 Commonly used methods

Although the presented studies addressed a variety of user-related topics in different experimental settings, some methods to assess the user-related concepts were used across studies (see Table 2.3). These methods are the L3Pilot questionnaire which was developed within the project (see Metz et al. 2020) to investigate relevant driver-related concepts, the van der Laan-scale (Van Der Laan, Heino, & De Waard, 1997) to asses usefulness and satisfaction of the ADF and the Karolinska-Sleepiness scale (Åkerstedt & Gillberg, 1990) to assess driver fatigue.

Table 2.3: Methods used in different supplementary studies. WoZ = Wizard of Oz study, DS = driving simulator study.

Method	Study no & short title							
	1	2	3	4	5	6	7	8
	WoZ long-term	DS long-term	On-road long-term	DS short-term	DS ambient light	WoZ take-over	Impairmens	Survey
L3Pilot questionnaire	X	X	X					
Van der Laan Scale	X	X			X			X
Karolinska Sleepiness Scale		X				X	X	

The L3Pilot questionnaire (see Metz et al., 2020) was developed within the project and covers the common RQs. Thefore it is suited to adress all common RQs within one study. In one part of the questionnaire, there is a list of statements with which drivers can agree or disagree on a 5-point Likert-scale. These items link directly to the RQs. Table 2.4 lists the different items and the related RQs.

Table 2.4: Items of L3Pilot questionnaire and their relation to common research questions.

L3Pilot Research Question	RQ Level 2	Questionnaire items
RQ-U1	Are drivers willing to use an ADF?	I would use this system if it was in my car.
		I would use the system during my everyday trips.
RQ-U3	What is the perceived safety of the ADF?	I felt safe when driving with the system active.
	What is the perceived comfort of the ADF?	Driving with the system active was comfortable.
	What is the perceived reliability of the ADF?	Sometimes the system behaved unexpectedly.
		The system worked as it should work.
		The system acted appropriately in all situations.
	What is the perceived trust of the ADF?	I trust the system to drive.
RQ-U5	What is the effect of ADF use on drivers' level of stress?	Driving with the system was stressful.
	What is drivers' level of fatigue while using the ADF?	Driving with the function on long journeys would make me tired.
	What is drivers' workload while using the ADF?	Driving with this system was demanding.
RQ-U6	What is the effect of ADF use on driver attention to the road / other road users?	During driving with the system active, I monitored the surrounding environment more than in manual driving.
		I would want to monitor the system's performance.
	What is the drivers' risk perception while using the ADF?	During driving with the system active, I was more aware of hazards in the surrounding environment than in manual driving.
RQ-U9	What secondary tasks do or would drivers engage in during ADF use?	
	What is the frequency and duration of drivers' secondary task engagement during ADF use?	I would use the time the system was active to do other activities

3 Background Literature Review

One aim of L3Pilot is to test the viability of automated driving as a safe and efficient means of transportation. User acceptance and user behaviour are key factors in the success of automated driving in the market. Public acceptance of L3 automated driving was one focus of the international online surveys conducted in L3Pilot (reference to D7.1). User acceptance and behaviour, however, are likely to change after exposure to automated driving. Behavioural adaptation, the "behaviours which may occur following the introduction of changes to the road-vehicle-user system and which were not intended by the initiators of the change" (OECD, 1990) are observed for drivers using ADAS (Naujoks & Totzke, 2014; Rudin-Brown & Parker, 2004) and attitudes change with extended usage of L2 automated driving functions (ADF) (Dikmen & Burns, 2017). In L3 automated driving where the role of the driver changes from actively monitoring the driving task to remaining receptive to warnings, behavioural adaptation can be expected on different levels. This literature review will give on overview of the state-of-the-art of various user/driver-related topics around AD.

Trust in the automation is considered to be a key premise for the use and acceptance of AD (Kyriakidis et al., 2017). The initial attitudes towards automated driving or the 'acceptability' of AD is likely to change when drivers actually experience it. If drivers are subjected to critical situations when using the system, trust and thus their acceptance of the ADF decreases (Gold, Körber, Hohenberger, Lechner, & Bengler, 2015). Therefore, the repeated experience of system boundaries can affect the drivers trust and acceptance of AD.

Reducing the driver's responsibilities to monitor the driving scene offers a wide range of new **driver states**: Drivers are free to engage in secondary activities ranging from smartphone usage to watching movies or engaging in work-related activities. These opportunities for distraction however, might compromise the drivers' ability to take over vehicle control at system boundaries (Zhang, de Winter, Varotto, Happee, & Martens, 2019). Empirical evidence also suggests that automated driving promotes fatigue in drivers due to mental underload and boredom (Matthews, Neubauer, Saxby, Wohleber, & Lin, 2019; Vogelpohl, Kühn, Hummel, & Vollrath, 2019). Even though these driver states are clearly not allowed, drivers might be more prone to drive under the influence of alcohol or when heavily sleep-deprived. These ranges of driver states need to be considered in the design of automated vehicles.

Especially the transitions from automated to manual driving at system boundaries pose a challenge when the driver is disengaged from the driving task. **Take-over behaviour** was found to be impaired when drivers are engaged in secondary activities (Zhang et al., 2019), when they are fatigued (Vogelpohl et al., 2019) or under the influence of alcohol (Wiedemann et al., 2018).

Driver behaviour and attitudes towards automation are likely to change over time with increasing experience with the ADF. Martens and Jenssen (2012) suggest changes in driver behaviour on different levels: perceptive, cognitive, performance, driver state and attitudinal changes as well as changes in the adaptation to environmental conditions. The scope of the literature review will be driver behaviour and attitudes in automated driving and **possible changes with increasing experience**.

3.1 Behavioural Adaptation

Various models attempt to explain changes in driver behaviour that occur after the introduction of changes to the road-traffic system. These changes in driver behaviour are referred to as ‘behavioural adaptation’ (BA) which means the “collection of unintended behaviours that follows the introduction of changes to the road transport system” (preface, Rudin-Brown & Jameson, 2021). Behavioural adaptation has been observed as a consequence of road transport interventions such as the introduction of speed humps which resulted in uneven driving behaviour with stronger decelerations and accelerations or the introduction of public lightning on motorways which led drivers to increase travel speed (van der Horst, 2012). BA was also observed when drivers used ADAS such a ACC or Intelligent Speed Adapter (Rudin-Brown & Parker, 2004; Naujoks & Totzke, 2014). Not much is known however, about possible BA to ADS.

Early theories to explain changes in driver behaviour in road traffic research have a strong focus on risk perception and risk homeostasis. Taylor (1964) and later Näätänen and Summala (1974) explain changes in driver behaviour following traffic safety measures by the driver’s subjective risk monitor. The driver experiences a certain degree of subjective risk or fear in a traffic situation which drives their decisions and actions. Wilde (1982) took up the idea that driver behaviour is shaped by the driver’s subjective risk perception and developed the Risk Homeostasis Theory to explain changes in driver behaviour. According to Wilde, drivers have a target level of risk that they are willing to accept. If a change in the road traffic system increases or decreases the subjective level of risk, the drivers will adjust their behaviour such that their target level of risk is met. Empirical evidence for theories of risk compensation or risk homeostasis comes from several studies: One study on BA to antilock brake systems, a system designed to reduce accidents, found that drivers kept shorter distances to vehicles in front when driving a car with the antilock brake system (Sagberg, Fosser, & Sætermo, 1997). Several studies on the effects of seatbelt use on driver behaviour showed that when using a seat belt, drivers drove faster on average (Wilde, 2012). These observations support the assumption that drivers aim to reach their ‘target level’ of risk. The introduction of a measure that is assumed to increase safety and decrease the risk of having an accident leads drivers to riskier behaviour.

Rudin-Brown and Parker (2004) include trust as an important factor in their qualitative model of behavioural adaptation to Advanced Driver Assistance Systems (ADAS). Another key factor of their model is the driver’s mental model of the driving task. Drivers were found to change their behaviour when using ADAS in ways that were unintended by the designers of the systems. When using Adaptive Cruise Control (ACC), for instance, a system that keeps a set speed as well as a set distance to a front vehicle, drivers changed other parameters of driving behaviour. They engaged more in a secondary task when using ACC, paid less attention to the driving task and reacted more slowly to a hazardous situation. The drivers’ behavioural adaptations were associated with the personality variables sensation seeking and locus of control. Drivers with an external locus of control reacted slower to a hazard situation than drivers with an internal locus of control. High sensation seekers also had more lane position variability when using ACC than low sensation seekers. BA was observed not only for ADAS, but also for warning systems: When using a congestion tail warning, drivers were found to engage more in a secondary task, drove with higher speeds and decreased the distance to a vehicle in front (Naujoks & Totzke, 2014).

Studies on behavioural adaptation to ADAS usually assessed drivers' behaviour in terms of driving parameters such as time headway or mean speed (Naujoks & Totzke, 2014; Sagberg et al., 1997) or the degree of distraction (Naujoks & Totzke, 2014; Rudin-Brown & Parker, 2004). These parameters are not applicable when investigating the behavioural adaptation to higher automated driving. To assess drivers' behavioural changes that were not anticipated by the designers, a definition of expected driver behaviour is needed. The SAE taxonomy (SAE, 2021) defines the role of the driver when using L3 ADF as such: "[…] with the expectation that the DDT fallback-ready user is receptive to ADF-issued requests to intervene, as well as to DDT performance-relevant system failures in other vehicle systems, and will respond appropriately" (p. 31).

In a driving simulator study, Jamson, Merat, Carsten, and Lai (2013) investigated changes in driver behaviour when using a highly automated driving system compared to manual driving. In the automated driving condition, drivers' arousal was reduced resulting in a higher percentage of eyelid closures. The use of in-vehicle entertainment was higher during automated driving, and drivers executed fewer lane changes and spent more time in the middle lane when driving in automated mode which resulted in a longer journey time.

When investigating BA to ADFs an approach needs to be chosen that considers the fact that in highly automated driving there is a fundamental change in the driver's role. The classic method of comparing certain driving parameters when using a system to driving without the system is not applicable when investigating BA to driving at higher automation levels. Martens and Jenssen (2012) generate categories of changes in drivers' behaviour due to ADAS. Those categories seem relevant when investigating BA to ADFs. Behavioural changes are defined in terms of:

- Perceptive changes (seeing, hearing, feeling)
- Cognitive changes (comprehending, interpreting, prioritising, selecting, deciding)
- Performance changes (driving, system handling, error)
- Driver state changes (attentiveness/awareness, workload, stress, drowsiness)
- Attitudinal changes (acceptance, rejection, overreliance, mistrust)
- Changes in the adaptation to environmental conditions (weather, visibility, etc.)

It can be hypothesised that behavioural changes due to an ADF are interconnected. An increase in trust in the ADF, for instance, may lead to a higher willingness to engage in secondary activities, which could then lead to a decreased perception of the environment. Such links must be considered in the assessment of BA to ADFs.

Other models of BA take driver variables into account and consider the impact of the driver's personality variables, such as sensation seeking, locus of control, or the driver's (preferred) driving style (Martens & Jenssen, 2012) or the situational context (Saad, 2006).

One aim of the supplementary studies on driver behaviour and acceptance in L3Pilot was to investigate long-term effects of AD on driver behaviour and possible BA over time. Therefore, in three studies, drivers used an ADS repeatedly on several occasions. Changes in acceptance and trust, as well as driver behaviour in take-over scenarios and driver state were investigated in these studies.

3.2 Transitions from Automated Driving to Manual Driving

When the ADF is operating in automated mode, the driver is not responsible for any part of the driving task. Driver actions are only required at system boundaries. A take-over request is issued to the driver prompting him to take back vehicle control and drive manually. A challenge of these take-over situations is described by Bainbridge (1983) as an “irony of automation”: In regular, less demanding situations, the automation performs the task, and the human operator or the driver is only required in complex, potentially critical situations that the automation cannot handle. Usually, a certain time budget is provided for drivers before the system boundary is reached. The driver’s ability to take back vehicle control and handle the take-over manoeuver, depends, among other factors, on the modality and intensity of the TOR and the complexity of the driving situation (Naujoks, Mai, & Neukum, 2014; Wiedemann et al., 2018), as well as the driver state and the take-over time budget (Zhang et al., 2019).

When drivers gain more experience with take-over requests, their take-over performance is likely to change. When drivers experienced a second take-over request, the mean take-over time was shorter than for the first take-over (Zhang et al., 2019). This finding suggests that there is some kind of “learning effect” for take-over situations.

3.3 User Acceptance and Trust

A driver’s acceptance of AD is, besides technological readiness, one of the most crucial factors for user acceptance and successful deployment (Zhang, Tao, Qu, Zhang, Lin & Zhang, 2019). Most people probably already have certain expectations or attitudes towards automated vehicles without having experienced them in real life. This is referred to as acceptability, a prospective judgement about such systems. Acceptance, in contrast, describes attitudes towards the system after having experienced it. Acceptability does not necessarily lead to acceptance after using the system, and conversely, a lack of acceptability before the first encounter does not necessarily mean that users will reject the system after experiencing it (Jamson, 2013).

A number of surveys has been conducted to assess public acceptance of AD: In a 2012 survey with 17,400 vehicle owners in the United States, 37% showed interest in buying a vehicle with AD capacity (Power, 2012). Another survey with nearly 5000 repondents from 109 countries asked about their attitudes towards AD and found that AD was rated a easier than manual driving, but also as less enjoyable (Kyriakidis, Happee, & de Winter, 2015). The higher the automation level, the more willing drivers were to engage in NDRAs. One fifth of the respondents was not willing to pay more than $0 for an ADS. The respondents were most concerned about software hacking, legal issues and safety.

Especially the (perceived) safety of an ADS is a key factor for its safe usage. If drivers do not perceive the system as safe, they will not trust it. And if drivers do not trust the automation they will not use it (disuse; Parasuraman & Riley, 1997). On the other hand, if drivers over-rely on the automated system, this might lead to decision errors, for example, in terms of not responding appropriately to TORs. Positive effects of increasing acceptance of automated vehicles can be found already after the first drive. Older drivers are found to report higher trust levels than younger drivers. Drivers who have experienced crashes or safety-critical situations report lower trust levels (Gold et al., 2015). Trust is closely tied to the perceived reliability of an automated system. If the perceived reliability increases, trust is likely to increase as well. In a survey of 109 users of Tesla's Autopilot (a combination of automated longitudinal and lateral vehicle control, SAE L2) conducted by Dikmen and Burns (2017), initial trust (referring to acceptability) was compared to the level of trust after a certain period of use. Trust levels increased clearly. Trust in the system was positively correlated with frequency of use, knowledge about the system, ease of use, and perceived usefulness of the Human-Machine-Interface (HMI). The authors hypothesized that the increase in trust over time is related to the fact that drivers become more comfortable with the system. The introduction to an ADS is a key determinant for the building of trust in the system (Beggiato & Krems, 2013; Körber, Baseler, & Bengler, 2018). It is reasoned that a realistic description of the system capabilities is a precondition for building trust in the automation. Even automation failures did not affect trust negatively if drivers were warned beforehand that they might occur (Beggiato & Krems, 2013). However, an unreasonably high level of trust can lead to drivers neglecting their monitoring duties or poor take-over performance. In their internal testing of an SAE L3 system, Waymo (2018) found that due to over-trust in the technology, human drivers were not carefully monitoring the system and environment and were not able to safely take control when needed.

Driving comfort is also considered to play a part in general acceptance. Driving comfort is highly subjective, differing between individuals and affected by physical, physiological, and psychological factors. It results from the interaction between an individual and the environment. In the context of AD, the implemented driving style of the vehicle is a key aspect of driving comfort. Accelerations and sudden movements should be as minimal as possible (Bellem, Thiel, Schrauf, & Krems, 2018). Drivers consider driving styles that are similar to their own driving style to be comfortable (Hartwich, Beggiato, & Krems, 2018).

The acceptance of AD is also highly related to its perceived usefulness. The perceived usefulness of an ADF for the user might increase with an increasing automation level. When the driver is not required to monitor the system's performance and is allowed to engage in other activities they will perceive the system as more useful. Several surveys have been conducted on the activities drivers want to engage in while driving in automated mode. The perceived usefulness of the AD depends on the extent to which drivers are able to perform these activities (Naujoks, Wiedemann, & Schömig, 2017).

Trust is a key variable for the use of AD (Lee & See, 2004). The driver's level of trust in the ADS not only influences the overall usage, it only influences the driver's state while using the system: Studies suggest that high trust in the ADS is linked to higher levels of drowsiness (Kundinger, Wintersberger, & Riener, 2019) and higher engagement in NDRAs (Kyriakidis, Happee, & de Winter, 2015). Empirical evidence of the effects of AD on the driver state are presented in the next section.

3.4 Driver State in Automated Driving

Relieving the driver from vehicle control in L3 AD introduces a variety of changes in driver state. The most obvious change is that the driver can engage in secondary activities and might therefore be cognitively or motorically distracted. This distraction can cause drivers to be “out of the loop”, which is linked to two major issues: loss of manual driving skills and loss of awareness of the state and processes of the system (Endsley & Kiris, 1995; Merat et al., 2018). Within a short time frame after a TOR by the system the driver needs to regain situation awareness (Endsley, 1995) and get back into the loop in order to be able to resolve the situation safely. Situation awareness is “the perception of environmental elements and events with respect to time or space, the comprehension of their meaning, and the projection of their status after some variable has changed, such as time, or some other variable, such as a predetermined event” (Endsley, 1995). Especially when drivers are engaged in other activities, these issues might be of concern. When drivers are not actively engaged in the driving task they might lose their comprehension of the environment and thus not react appropriately when required to retake vehicle control.

In highly automated driving, the drivers’ mental workload is lower than in manual driving or driving with ACC (De Winter, Happee, Martens, & Stanton, 2014) which can be seen as a major benefit of AD for drivers. While the reduction in mental workload is a benefit for the drivers’ comfort, mental underload due to automation, may result in an increase in drowsiness or fatigue (Greenlee, DeLucia, & Newton, 2018; Schömig, Hargutt, Neukum, Petermann-Stock, & Othersen, 2015; Vogelpohl et al., 2019). However, the engagement in secondary activities has the potential to counter fatigue by activating the driver (Naujoks, Höfling, Purucker, & Zeeb, 2018; Neubauer, Matthews, & Saxby, 2014). Driver drowsiness can result in longer take-over times (Vogelpohl et al., 2019) and drivers falling asleep (Omae, Hashimoto, Sugamoto, & Shimizu, 2004) which then causes a major safety issue.

An additional concern with the introduction of automated driving is that drivers might misuse the systems and use them in ways they are not designed to be used. A study on the usage of SAE L2 systems found that in 57% of all safety-critical events, drivers had misused the system, for example by engaging in secondary tasks, driving with hands off the wheel or using the system on roads it was not designed for (Kim, Song, & Doerzaph, 2020). In L3 ADF, one potential misuse of automation is that drivers use the system when they are not fully fit to drive themselves, e.g., when they are sleep deprived or drunk. This misuse can have serious safety consequences: At a blood alcohol level of 0.08%, it took drivers longer to respond to a TOR, i.e. to put their hands on the wheel and deactivate the ADF, and their driving behaviour was impaired (Wiedemann et al., 2018). Mean reaction times to a take-over request were also extended when drivers were drowsy (Naujoks et al., 2018).

3.5 User Experiences from a Long-term Perspective

Studies investigating user experiences of ADAS and ADFs mostly assess the drivers' behaviour and attitudes when they first encounter the new technology. In most studies, for practical reasons, only the first 1–2 hours of using a new technology are investigated. However, it is obvious that after a certain time of using and experiencing the behaviour of the system in several use cases, the drivers will adopt their behaviour accordingly. However, long-term usage is assessed very rarely since this is rather complex and expensive. The EU-funded AIDE project is one of the rare examples that has focused on the long-term effects of using ADAS. Project findings showed that with increased exposure drivers overrode an Intelligent Speed Adapter more frequently, with half of the drivers experiencing a "change point" in how they used the system. Initially, after the introduction of the ISA, drivers' speeding decreased to a lower level but later increased again steadily. In contrast, for drivers who had a Cruise Control implemented in their car, the proportion of "non-users" decreased over time. Drivers using a combination of Forward Collision Warning + Lane Departure Warning drove at shorter headways the longer they used it. It is assumed that the behavioural changes depend on the specific functionality of the respective system (Portouli et al., 2006).

The "learning process" is crucial for drivers to gain an appropriate understanding of the system's functionality as well as system limits and helps to build an appropriate level of trust. It is emphasised that this learning process will take some time and requires experience of the system in different situations and different environments. Two phases in the learning process are suggested: in the "learning phase" the driver learns how to operate the system, identifies system limits, and internalizes the system functionality. The learning phase heavily depends on the way the system is introduced to the driver. At the second stage, the "integration phase", the driver integrates the system into the management of the overall driving task through increasing experience in different situations (Saad et al., 2004).

When testing ADAS in the AIDE project, the focus was on directly observable behavioural changes among the drivers due to the ADAS, mainly in terms of changes in driving parameters. However, when assessing L3 vehicles, the approach must be adapted. Since the vehicle is guided by the automated system, most of the time changes in human driving behaviour can only be assessed to a limited extent. However, attitudes towards the automation can change dramatically over time, for instance when experiencing the system in different traffic situations.

When investigating long-term effects in user behaviour and experience, one question is: How long is long-term? Martens and Jenssen (2012) define five phases of BA to ADAS with defined durations:

- First encounter: First day (1–6 hours)
- Learning: 3–4 weeks
- Trust: 1–6 months
- Adjustment: 6–12 months

- Readjustment: 1–2 years

For highly automated driving systems a more rapid adaptation process might be expected. When using ADAS, drivers occasionally experience system actions that only affect parts of the driving task while they are otherwise driving manually. In highly automated driving, however, the driver's role changes fundamentally, removing them from the driving task completely. Therefore it can be expected that the adaptation process will be faster for AD than for ADAS. It is emphasized that not only the length of experience with a system affects BA but also the experience of the system in different situations.

The *First encounter* phase depends greatly on how intuitive and self-explaining the HMI is. The *Learning* phase still depends highly on the HMI, especially in terms of required system input. The *Trust* phase is mainly characterised by a shift in locus of control (Ajzen, 2002) from the driver to the vehicle. Related problems might be overreliance, passivity, and drowsiness. In the phases *Adjustment* and *Readjustment*, drivers adjust their adapted behaviour depending on their experience of (critical) situations and system limitations. It can be expected that trust plays an important role in the BA to AD, and indeed, for the overall acceptance of the system. According to Muir (1987), trust depends on the degree of experience with automation and thus can be expected to change over time.

One study investigated secondary task engagement during highly automated driving from a long-term perspective. Large, Burnett, Morris, Muthumani, and Matthias (2017) invited six drivers to undertake five 30-minute journeys with a highly automated system in a driving simulator. They were encouraged to use the system just as they would in a real automated vehicle. Participants were asked to bring with them any objects or devices that they would be willing to engage with during the drives. The most common activities during the drives were reading articles or magazines, using mobile devices for social networking activities, web browsing, and watching programmes or films on a laptop. Unfortunately, no findings on changes in behaviour over time were reported.

4 Wizard of Oz Study on long-term behavioural adaptation

In this chapter an on-road study is described that explored the change of acceptance and usage of an L3-ADF with growing experience using a Wizard of Oz vehicle. The work was conducted by the Federal Highway Research Institute (BASt).

4.1 Aim and research questions

The presented Wizard of Oz study aims to answer research questions from the L3Pilot research question list and additional RQs developed by BASt:

- RQ-U1: Are drivers willing to use an ADF?
- RQ-U3: What is the perceived trust of the ADF?
- RQ-U9: What secondary tasks do drivers engage in during ADF use? What is the frequency and duration of drivers' secondary task engagement during ADF use?
- RQ-U10: How do drivers respond when they are required to retake control in expected use cases? How do drivers respond when they are required to retake control in unexpected use cases?
- RQ-U11: How often and under which circumstances do drivers choose to activate/deactivate the ADF?
- BASt-RQ 1 How does users' trust and acceptance develop if they use an ADF several times in real traffic?
- BASt-RQ 2 How does a non-safety-critical disturbance of the ADF affect the development of trust and acceptance?

Both BASt-RQs specifically address multiple ADF use in real traffic. The hypotheses for the BASt-RQs are:

- BASt-RQ 1: Trust and acceptance will increase with increasing driving experience.
- BASt-RQ 2: The non-safety-critical system disturbances inhibit the development of trust and acceptance, so that the level of trust and acceptance of the respective drivers remains on a lower level than that of the drivers with a well-functioning system.

4.2 Methods

4.2.1 The Wizard of Oz approach

The so-called "Wizard of Oz" principle can be used to simulate an ADF in a research vehicle. There are two drivers on-board of the vehicle, the test participant who controls the vehicle during periods of manual driving, and a second driver who controls the vehicle during all periods of allegedly activated ADF. As the second driver is hidden, the test participant believes that an automated driving system is activated during periods of automated driving. This technique is especially useful for studies on human machine interaction and psychological issues such as trust or acceptance.

4.2.2 BASt research vehicle and HMI

The BASt Wizard of Oz research vehicle is based on a Volkswagen Caddy Maxi. There are three persons present in the vehicle: The participant sits in the regular driver's seat; the second driver, the trained Wizard of Oz driver, is seated in the second row of seats behind a one-way-pane and thus, invisible to the participant. The examiner sits behind the second driver in the third row of seats. This driver is introduced as an engineer to the participant.

Since there is no real technical function, many common limitations of real automated research vehicles do not apply. Driving in various traffic densities, with velocities of up to 130 kph, and performing "automated" lane changes is possible. The participant does not have to monitor the alleged Level 3 system, like a safety driver would have to. Therefore, ordinary drivers can be chosen as participants for studies.

The German Federal Highway Research Institute was granted permission to perform studies with its Wizard of Oz vehicle in real traffic, including the transition of control from the participant to the second driver and back while driving. The participant is allowed to perform non-driving related activities (NDRAs, e.g. texting, reading) when being driven by the second driver.

Figure 4.1: BASt research vehicle.

In order to activate and deactivate the alleged ADF and inform the participant about the current system status, an HMI was added to the vehicle (see Figure 4.2). By pressing a single button on the steering wheel, the participant can decide whether to drive manually or use the automated function. The current system status (system not available, system available, system active, TOR) is shown on a screen next to the speedometer. Changes in system status are also indicated by signal tones. Additionally, a tablet PC is attached to the centre console of the dashboard. It can be used by the participant to perform NDRAs during automated driving, such as games etc.

Figure 4.2: Participant's view in BASt's research vehicle.

4.2.3 ADF used in study

With the alleged automated driving function of the Wizard of Oz vehicle, a motorway chauffeur (SAE Level 3) was simulated. This included a traffic jam chauffeur, automated lane changes, and velocities from 0 up to 130 kph. In addition, a minimal risk manoeuvre (MRM) was available, performed by the second driver, in case the participant did not respond to a take-over request. System limitations were rain, snow, ice, fog, poor lighting conditions, road works and motorway entries and exits. Hitting any system limitation caused a TOR of the alleged automated driving function. The TORs were set off by the second driver.

4.2.4 Study design and environment

In order to observe the development of trust and acceptance over time, each participant used the ADF three times with intervals of one week between each drive. The experimental design was a between-subjects design with two groups: It had been planned to assign 15 participants to each group, but due to the Covid pandemic the study had to be curtailed to six participants in total. Group 1 (control group) did not experience any planned disturbance and used a smoothly operating automated driving system. Group 2 (treatment group) participants experienced three TORs due to an alleged system disturbance during the second drive (see BASt-RQ 2): With the TOR issued through the HMI, a female voice stated “System disturbance, please take over!” and a permanent warning sound was played until the participant deactivated the ADF. As for the normal TORs, the time budget for the system disturbance TORs was 10 seconds. After each disturbance, the system was not available for a short period of time which varied between one to two minutes. Subsequently, participants were able to switch the automated driving mode on again. If the automated driving mode was activated, the next disturbance followed after approximately five minutes. The plan of the study design in Table 4.1 provides further information.

Table 4.1: *Experimental design used for the Wizard of Oz study.*

	Group 1 (control group)	Group 2 (treatment group)
Short test drive	Familiarisation with vehicle, ADF, & transition	Familiarisation with vehicle, ADF, & transition
1st drive	No planned disturbance of automated driving – only TORs before leaving motorway and when otherwise necessary in case of fortuitous events	No planned disturbance of automated driving – only TORs before leaving motorway and when otherwise necessary in case of fortuitous events
2nd drive	No planned disturbance of automated driving – only TORs before leaving motorway and when otherwise necessary in case of fortuitous events	**Three planned disturbances of automated driving (TORs) in short succession with subsequent short periods of manual driving** plus: TORs before leaving motorway and when otherwise necessary in case of fortuitous events
3rd drive	No planned disturbance of automated driving – only TORs before leaving motorway and when otherwise necessary in case of fortuitous events	No planned disturbance of automated driving – only TORs before leaving motorway and when otherwise necessary in case of fortuitous events

Disturbances during the second drive were planned based on the following rationale: The first drive did not differ between the groups. It was therefore expected that the level of trust and acceptance should on average be the same for both groups after the first drive. The disturbance set to occur during the first half of the second drive allowed participants to show a potentially changed behaviour during the second half of the second drive, right after the disturbances. The third drive again did not differ between both groups. It allowed further development of trust and acceptance, based on the participants' experience during the first and second drive.

Because the study took place in real traffic, unplanned but necessary TORs occurred during the drives in addition to the planned TORs. Road works opened and closed during the testing periods and weather conditions changed, so that participants' experience of the study environment somewhat differed. The planned TORs took place at standardised locations, but they might have taken place at slightly different places and times during the drives due to traffic situations.

The study was conducted on a German motorway. A hilly and curving section of Motorway A4, from the "Frankenforst" junction to the "Eckenhagen" junction and back, was chosen: with a total of 96 km (approx. 1 hour of driving time), the route combined dense traffic in the western part (near Cologne) and fairly light traffic in the eastern part (countryside), two and three lanes per direction, and various speed limits from 80 kph up to no speed limit. In case of no speed limit, the participants were instructed to drive at a maximum speed of 130 kph.

The duration of one test slot did not exceed two hours; therefore, so the total time spent per participant was a maximum of six hours. Test drives were executed two times a day (at 10 am and 2 pm) from Monday to Friday. The three test drives of each participant were scheduled at an interval of one week and were conducted on the same weekday and time of the day, e.g. on Wednesdays at 10 am.

4.2.5 Overall instructions and study procedure

4.2.5.1 Overall instructions

Before the first drive, the participant received detailed instructions on how to operate the ADF. Each system status of the ADF was explained in detail. Training on how to correctly activate and deactivate the ADF took place in the parked vehicle, using a demonstration mode of the HMI. Furthermore, system limitations (e. g. bad weather conditions or road works) were explained to be a cause for a TOR. After that, the participant had to explain each system status to the examiner to make sure they had a correct understanding of the system.

The participants were asked to try out the ADF during the test drives, but were free to switch it off, if they did not feel comfortable. As long as the HMI indicated that the system was available, the participants were allowed to switch it on and off as they pleased. It was explained that when driving in automated mode, they were not responsible for controlling the vehicle but had to stay vigilant for the TOR. Before pressing the button to deactivate the ADF, they had to stop executing the NDRA, regain orientation and situational awareness by looking in the mirrors and outside the windshield, and then finally take over control of the vehicle. After that, they drove manually and were thus fully responsible for driving the vehicle. During automated driving, they did not need to monitor the system and were legally allowed to perform an NDRA. In order to observe the most natural behaviour of the participants while driving with the ADF engaged, a standardised NDRA was deliberately not used. Insetad, participants were allowed to use their own brought along electronic devices (e.g. smartphones) or magazines as NDRA. Additionally, several magazines and a fixed tablet with games were available on-board the vehicle. NDRAs that could potentially obstruct the second driver's view or hearing (e.g. newspapers, loud music, or video sound) and NDRAs that are difficult to interrupt or reduce the participants' ability to notice the TOR (e.g. phone calls or wearing headphones) were not allowed.

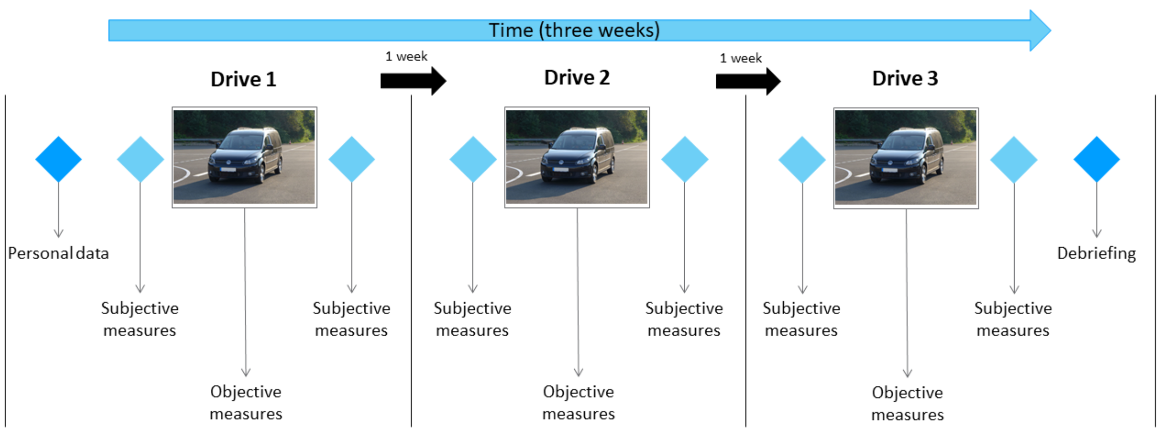

Figure 4.3: Study procedure.

4.2.5.2 First drive

The participants were welcomed in the lobby of the Federal Highway Research Institute (BASt) and completed the first questionnaire on technology affinity and their initial level of acceptance towards manual and automated vehicles. The results for acceptance towards automated driving are used as a baseline. Trust in automated vehicles was not collected at this stage, because no participant had driven a Level-3-vehicle before. For detailed information on the questionnaires, see chapter 4.3.2 “Subjective Data”. After completing the questionnaires, the participants were guided to the research vehicle, familiarised with the stationary vehicle; they practised the use of the HMI and received detailed instructions (see above). The eye tracking system was calibrated and the participants signed forms of consent for data recording and a confidentiality agreement and commited to adhere to traffic rules.

A short practice ride on the motorway followed to train the participants on how to activate and deactivate the ADF in a realistic setting. During a short break on a parking lot, participants completed another questionnaire with their first ratings of trust and acceptance towards the ADF that s/he had experienced a few minutes earlier. After that, the first test drive began.

After the drive, participants answered a questionnaire on their trust and acceptance resulting from experiences with the ADF during the first test drive. They received a compensation and were seen off.

4.2.5.3 Second drive

Before the drive, participants completed a questionnaire on trust and acceptance towards the ADF and received a refresher training on the vehicle’s HMI and their responsibilities. The treatment group experienced three non-safety-critical system disturbances of automated driving during the first half of the drive, whereas the control group experienced a well-functioning system without disturbances. After the drive, participants completed a questionnaire on their trust and acceptance towards the ADF. They received a compensation and were seen off.

4.2.5.4 Third drive

Before the drive, participants completed a questionnaire on trust and acceptance towards the ADF. After a short refresher on the vehicle’s HMI and participants’ responsibilities, the third drive began. There were no planned differences between the third drives of control and treatment group. After the drive, participants completed a questionnaire on their trust and acceptance towards the ADF, plus several additional questions such as their overall impressions on using the ADF and their perceived level of comfort. They received a compensation of 60 € and the Wizard of Oz principle was revealed. After they had the possibility to ask questions on the study, they were thanked for their participation and seen off.

4.3 Data Sources and Analysis

For the analysis of the data, IBM SPSS 25, Microsoft Excel, and ELAN 5.9 were used.

4.3.1 Objective data

The Wizard of Oz vehicle is equipped with several types of sensors and measuring systems. Basic driving data (3-axis acceleration, velocity, lane position, time headway) is recorded in order to describe the driving situation. Furthermore, the “driver in control” is logged, which allows take-over times and duration of automated and manual driving to be calculated and thus, the time portions of automated and manual driving as an indicator for acceptance and trust. A “Smart Eye” remote eye-tracking system is used to collect data on the participant’s gaze direction (e.g. the NDRA or monitoring gazes towards the road or instrument cluster) which can indicate participants’ trust. Several video cameras record participants’ behaviour in the vehicle as well as the surrounding traffic environment.

4.3.2 Subjective data

Before the first drive, information on basic demographics and participants’ affinity towards technology was collected. The questionnaire “Technikaffinität erfassen” (“Assessing technology affinity”) by Karrer, Glaser, Clemens, & Bruder (2009) was used. Furthermore, acceptance towards manual driving was assessed with the acceptance questionnaire by van der Laan, Heino, & de Waard (1997). It consists of two subscales: “usefulness” with five items and “satisfaction” with four items. The subscales are reported individually. **Please note**: In van der Laan’s questionnaire, negative numeric values stand for high acceptance and vice versa. For an easier readability of the results, all satisfaction and usefulness values were recorded so that positive numeric values always depict high acceptance and vice versa.

Before and after each drive, participants completed one questionnaire on trust and one on acceptance in order to trace their development over time. For acceptance, the acceptance questionnaire by van der Laan et al. was used (see above). Again, the subscales are reported individually. For trust, the questionnaire “Trust in Automation” by Körber (2018) was used. It contains 19 items, 15 of which were used for this study to calculate the reported trust score. The subscales “Familiarity” and “Intention of Developers” were not used, because these constructs were not relevant for this study.

4.4 Analysis

The development of trust and acceptance are the focus of the Wizard of Oz study. Gaze behaviour, NDRA engagement, system usage, and self-reported levels of trust and acceptance in the questionnaires were analysed with regard to their development over time. Emerging patterns were reported.

Due to the Covid pandemic, testing had to be ended after six participants had completed the study. The collected data was analysed case-by-case, since the number of participants is too small for inferential statistical analysis.

A case-by-case-analysis was performed to gather deeper insights in participants’ trust in and acceptance of the tested ADF. Each participant was analysed separately with the same structure of subchapters. These subchapters are explained hereafter.

The subchapters “ADF use, NDRA engagement and gaze direction” mainly consist of three diagrams per participant, one for each drive (see example figure, left). The inner layer of the diagrams shows the time portions of ADF engaged (light blue), available (not engaged) ADF (dark blue) and not available ADF (grey) per drive. For the time when the ADF was engaged, the NDRA engagement was investigated in the middle layer: green sections for smartphone use, yellow sections for reading a magazine, red sections for tablet use, grey sections for no NDRA use and violet sections for other NDRA that are specified in the respective descriptions. In these descriptions, it is also stated how often the participant used the NDRAs during this drive (called “session”). Lastly, the outer layer of the diagrams shows where the participant was looking when he performed a NDRA. Colours matching those of the NDRA mean that the participants were looking at the NDRA. Black sections stand for monitoring gazes and gazes outside the vehicle. Due to the eye-tracking system’s poor reliability, gaze directions were assessed manually by video coding.

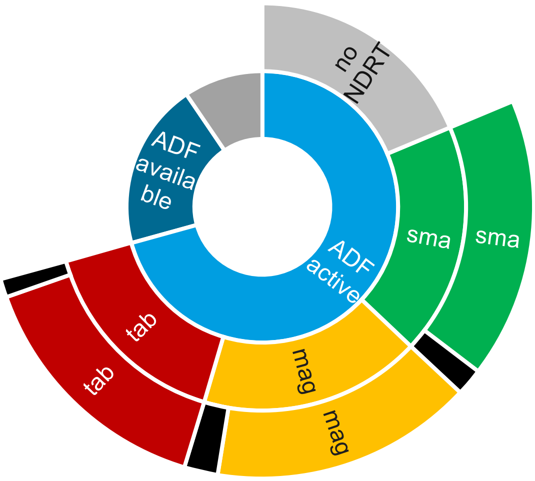

Figure 4.4: Example figure for ADF use, NDRT engagement and gaze direction

“Trust and acceptance over time” subchapters describe the development of each participant’s trust and acceptance over all three drives. For trust, higher numeric values indicate higher trust. Acceptance scores were recoded so that positive numeric values indicate higher acceptance and vice versa.

In the subchapter “Take-over behaviour” of each participant, charts show combined results of different aspects of system-initiated TORs. The figures show all TORs of all three drives with the take-over time in seconds: Each figure contains three curves and each of those curves stands for one of the three drives (in order). To depict circumstances of TORs and the participants NDRA engagement at the time the TOR was issued, shapes and colours were used as data points. This way, possible training effects and patterns can be seen easily.

Shapes depict the NDRA that the participant engaged in at the time of the TOR:

- ● no NDRA at TOR
- ■ reading magazine at TOR
- ▲ using smartphone at TOR
- ◆ using tablet at TOR
- ✖ performing office work at TOR

Colours indicate the reason for the system-initiated TOR:

- TOR b/o construction ahead
- TOR b/o motorway exit ahead
- TOR b/o system disturbance

4.5 Demographics

A total of six participants took part in the study. All of them were male and between 52 and 65 years old. Two participants were associated with the Federal Highway Research Institute (BASt): One was an employee of another department without knowledge about automated driving or the Wizard of Oz vehicle, the other was a family member of a BASt employee. None of the participants worked for a vehicle manufacturer, in automated driving development or as a test driver. The participants had vocational training (3/6) or a university degree (3/6).

All participants had held a driver's licence for between 18 to 45 years and had access to a car in their daily lifes. They drove approx. 18,000 km annually on average. Five participants stated they had cruise control or adaptive cruise control (ACC) in their own vehicle, but one of them reported not using it. Two participants had a lane-keep-assistant in their vehicle, but only one reported using it.

Before the study started, participants completed a questionnaire on technology affinity: With scores between 2.75 and 3.63 (3.34 on average), their technology affinity was in the medium range on a scale from 1 to 5. They rated their acceptance of manual driving between 0 and 1.8 on the usefulness scale (0.57 on average) and between -0.25 and 1.25 on the satisfaction scale (0.34 on average). Thus, participants rated their acceptance of manual driving neutral to positive on the inverted scale from -2 to +2.

4.6 Results

4.6.1 Participant A

Participant A was in the treatment group and experienced three TORs due to alleged system disturbances in the second drive.

With 96-98%, the participant's use of the ADF was very high in all three drives. While barely engaging in NDRAs in the first drive, a clear increase was seen in the second drive; enagement remained nearly constant in the third. The participant only engaged in smartphone use and reading magazines as NDRAs. Both, trust and acceptance were fairly high from the beginning. The take-over times ranged from 2.7 to 4.6 seconds without a clear trend in any direction; however, reading a magazine at the TOR seemed to prolong the take-over process. The three TORs due to an alleged system disturbance did not seem to influence trust, acceptance or take-over times.

4.6.1.1 ADF use, NDRA engagement and gaze direction (RQ-U1, RQ-U9)

The participant's use of the ADF was very high in all three drives. While barely engaging in NDRAs in the first drive, a clear increase was observed in the second drive; engagement remained nearly constant in the third. When using NDRAs, the participant mainly looked at the NDRAs too, but still with monitoring gazes.

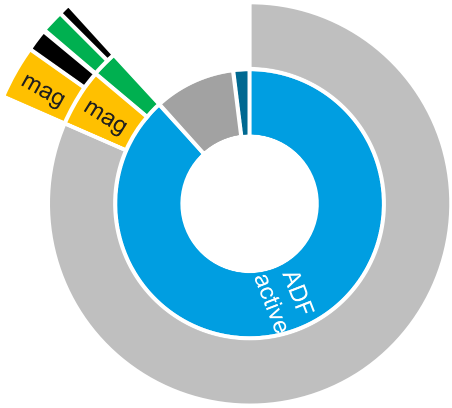

Figure 4.5: 1st drive of participant A

Participant A used the ADF nearly 98% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged comprised six activations with lengths between approx. 3 and 18 minutes (7 minutes on average). The participant spent only 2% of that time on smartphone use (4 sessions, green, middle layer) and 5%on reading a magazine (4 sessions, yellow), mainly looking at the respective NDRAs (outer layer, 30.3% monitoring gazes). For the remaining time, he did not perform a NDRA.

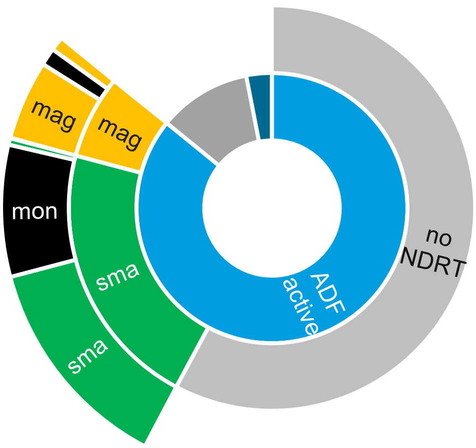

Figure 4.6: 2nd drive of participant A

Participant A used the ADF nearly 96% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged comprised eight activations with lengths between approx. 2.5 and 18 minutes (5 minutes on average). The participant spent about 25% of that time on smartphone use (10 sessions, green, middle layer) and 8% on reading a magazine (4 sessions, yellow), mainly looking at the respective NDRAs (outer layer, 31.9% monitoring gazes). For the remaining time, he did not perform a NDRA.

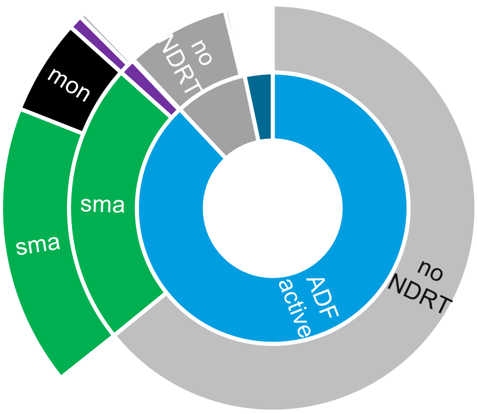

Figure 4.7: 3rd drive of participant A

Participant A used the ADF over 96% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged comprised six activations with lengths between approx. 3 and 18.5 minutes (7 minutes on average). The participant spent about 25% of that time on smartphone use (7 sessions, middle layer, green) and under 2% on using a smartphone and reading a magazine simultaneously (1 session, violet), mainly looking at the respective NDRAs (outer layer, 26.1% monitoring gazes). For the remainder of the time that theADFwas engaged, he did not perform a NDRA.

Participant A continued using smartphone and magazine in one case even after the take-over and spent a total of 13.5 seconds with distracted manual driving.

4.6.1.2 Trust and acceptance over time (RQ-U3, BASt-RQ 1, BASt- RQ 2)

The participant's self-reported trust level is fairly high, varies moderately and does not show a trend or signs of further adaptation to the system. After the second drive with three system disturbances, the trust level is slightly higher than before the second drive.

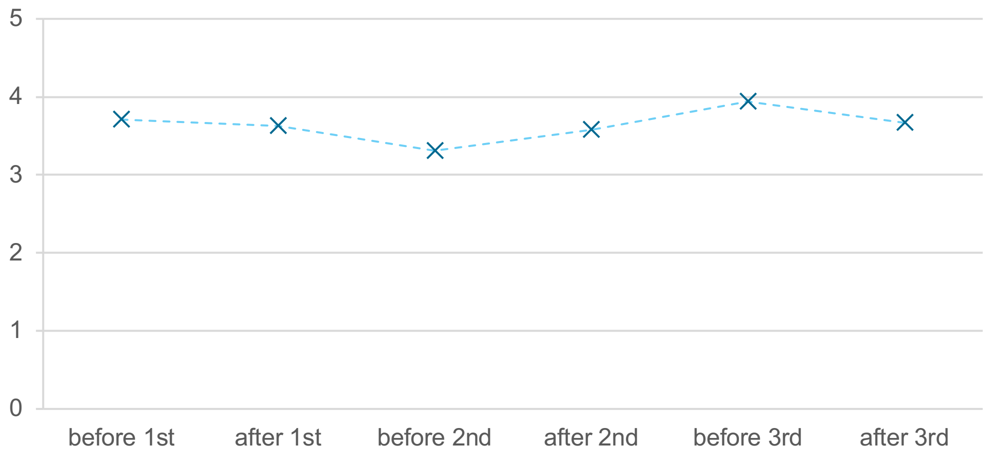

Figure 4.8: Development of trust over time (higher values indicate higher trust).

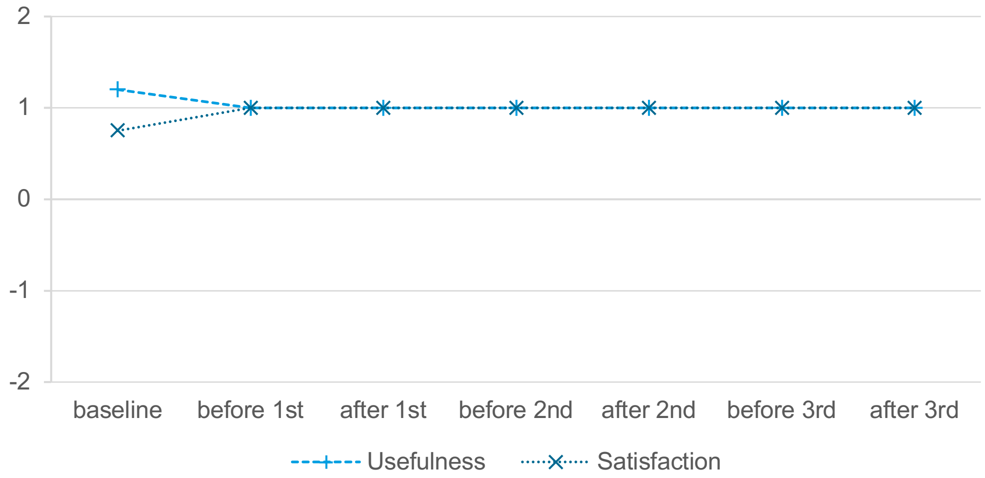

Figure 4.9: Acceptance rating scales (usefulness and satisfaction) over time (higher values indicate higher usefulness/satisfaction).

The levels of acceptance (usefulness and satisfaction scale) were consistently high in all three drives. Only the baseline measurement differed slightly. There were no changes in acceptance after the second drive with three system disturbances.

4.6.1.3 Take-over behaviour (RQ-U10, RQ-U11)

Across the three drives, the participant had to respond to 20 TORs in total. There were no apparent training effects regarding take-over time. When the participant was engaged in reading a magazine, it took him noticeably longer to take over than in scenarios with smartphone use or without NDRA. In one case (highlighted with a yellow asterisk in the chart), the participant used his smartphone and read a magazine at the same time, when a TOR was issued. The three TORs that were due to a system disturbance did not differ from the other take-over scenarios.

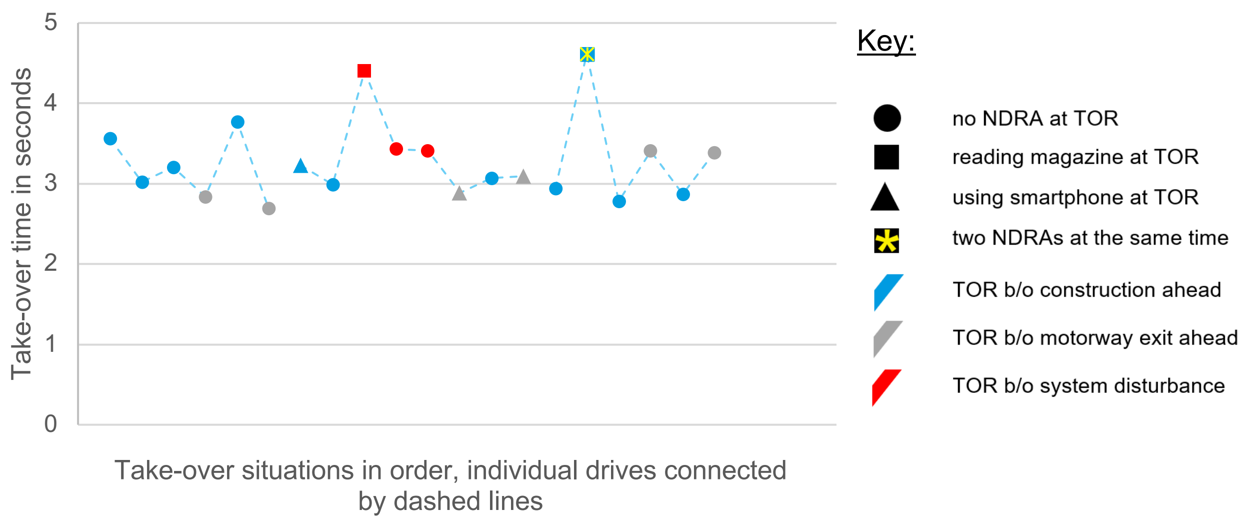

Figure 4.10: Take-over time in seconds over all three drives, NDRA engagement at TOR, reason for TOR.

Before deactivating the ADF, the participant always looked straight ahead to check traffic in front, but he looked in the rear-view mirrors in only six cases. In three cases, he did not finish his NDRA engagement before deactivation; one time, he kept using his smartphone in manual driving phase for approx. 20 seconds after take-over. Consequently, in 15 out of 20 take-over scenarios, his checking and take-over behaviour did not meet the instructions that had been given before each drive and could – depending on the traffic situation – be considered unsafe.

There were no participant-initiated deactivations of the ADF any drive.

4.6.2 Participant B

Participant B was in the control group and did not experience TORs caused by the alleged system disturbances.

The participant's use of the ADF was very high in all three drives (94-98%). While not engaging in NDRAs during the first drive, a clear increase was observed throughout the second and third drive. The participant only chose smartphone use as an NDRA and drank from a water bottle. His trust in the ADF was high at all measurement points, his satisfaction reached the maximum of the scale after the second drive. Take-over times ranged from 0.8 to 5.0 seconds with a trend to decrease over time. Reading a magazine at the TOR appeared to prolong the take-over process. Across all drives, the participant chose to deactivate the function upon his own request a total of nine times.

The participant showed signs of microsleep during the third drive. In the respective driving phases, he was not engaged in a NDRA other than drinking from a water bottle frequently. Near the end of the third drive, the wizard made a driving error: Due to unexpected road damage, the wizard swerved abruptly in the lane. The participant deactivated the ADF after the driving error, but activated the ADF quickly again.

In the debrief after the third drive, the participant reported that he was fully immersed into reading during the third drive and lost orientation as a result. In contrast, although he was also reading in the second drive, he reported he was always alert and did not really pay attention to the magazine.

4.6.2.1 ADF use, NDRA engagement and gaze direction (RQ-U1, RQ-U9)

The participant's use of the ADF was very high in all three drives. While not engaging in NDRAs in the first drive, a clear increase was observed in the second and third drive. When using NDRAs, he mainly looked at them, too, but still with monitoring gazes. Participant B was the only participant in this study who drank during the drives.

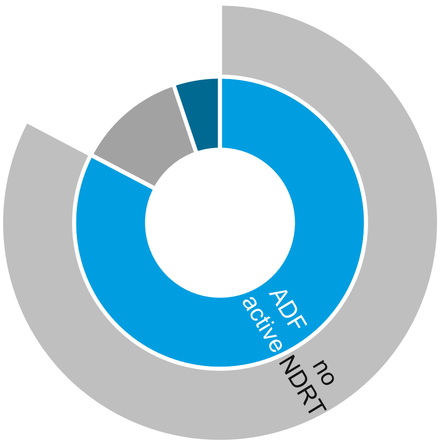

Figure 4.11: 1st drive of participant B

Participant B used the ADF 94% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged comprised eleven activations with lengths between approx. 19 seconds and 16.5 minutes (3.5 minutes on average). The participant never engaged in a NDRA; therefore, no gazes are reported (no third layer).

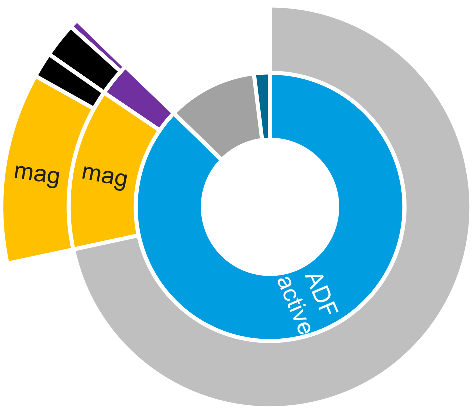

Figure 4.12: 2nd drive of participant B

Participant B used the ADF 98% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged comprised five activations with lengths between approx. 1 and 19 minutes (8 minutes on average). The participant spent about 15% of that time reading a magazine (1 session, yellow, middle layer) and 3% drinking from a bottle (5 sessions, violet). While reading the magazine, he mainly looked at the magazine (yellow, outer layer), but while drinking, he mainly looked outside the vehicle (black) (overall, 23.1% monitoring gazes). For the remaining time, he did not perform a NDRA.

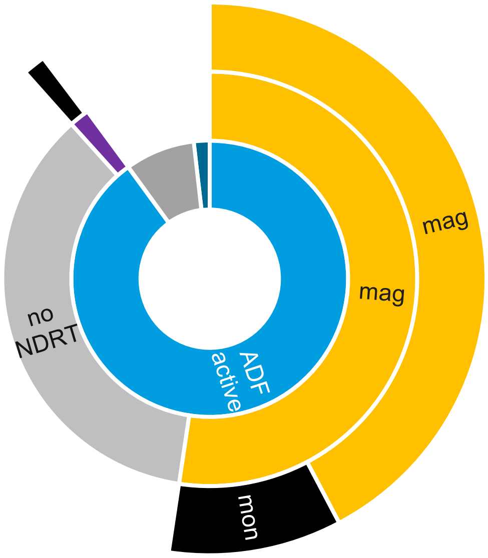

Figure 4.13: 3rd drive of participant B

Participant B used the ADF 98% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged comprised six activations with lengths between approx. 13 seconds and 18 minutes (7 minutes on average). The participant spent about 58% of that time reading a magazine (5 sessions, yellow, middle layer) and nearly 2% drinking from a bottle (3 sessions, violet). While reading the magazine, he mainly looked at the magazine (yellow, outer layer), but while drinking, he mainly looked outside the vehicle (black) (overall, 21.2% monitoring gazes). For the remaining time, he did not perform a NDRA.

4.6.2.2 Trust and acceptance over time (RQ-U3, BASt-RQ 1, BASt- RQ 2)

The participant's self-reported trust level was high overall, varied moderately and did not show a trend or signs of further adaptation to the system.

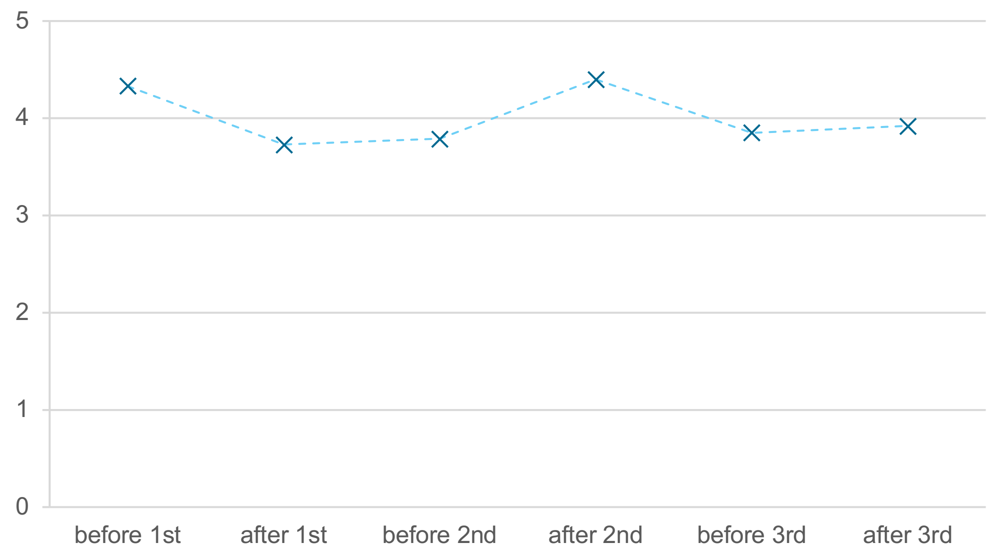

Figure 4.14: Development of trust over time (higher values indicate higher trust).

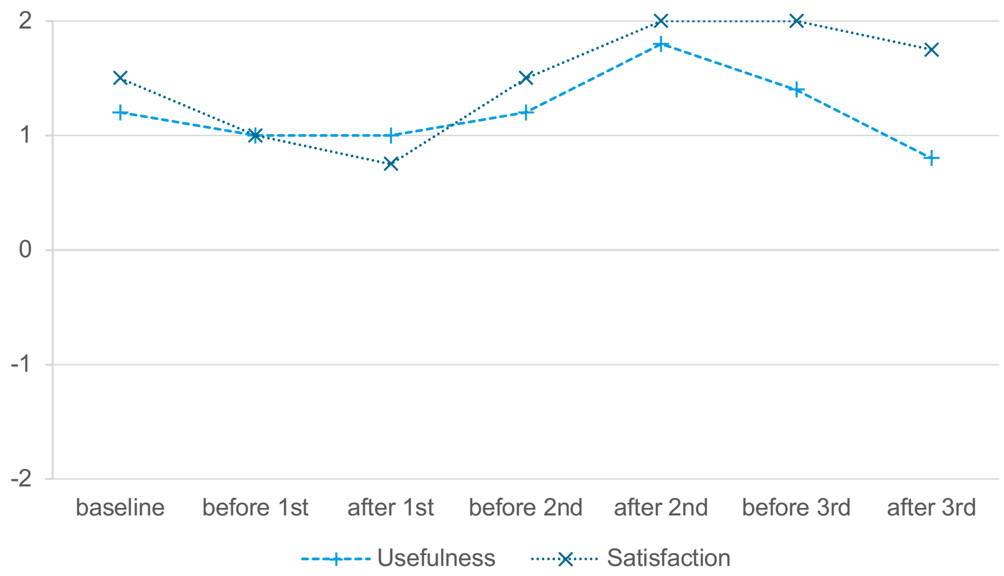

Figure 4.15: Acceptance rating scales (usefulness and satisfaction) over time (higher values indicate higher usefulness/satisfaction).

The levels of acceptance (usefulness and satisfaction scale) were high in all three drives, with a peak after the second drive. While usefulness was rated its lowest (but still relatively high) after the third drive, satisfaction reached a plateau at the maximal level (ceiling effect) after the second drive.

4.6.2.3 Take-over behaviour (RQ-U10, RQ-U11)

Across all three drives, the participant had to respond to 13 TORs in total. There were potential training effects for take-over time which in part could be explained by the predictability of TORs before motorway exits and construction sites. During the second TOR of the third drive, the participant was engaged in reading a magazine at a TOR: It took him noticeably longer to take over than in scenarios with no NDRA.

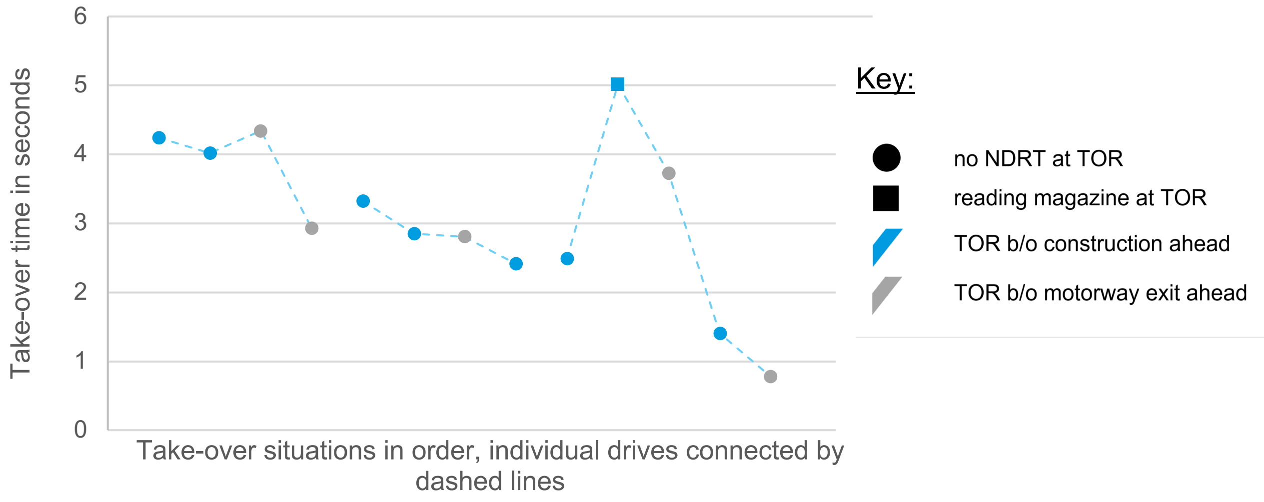

Figure 4.16: Take-over time in seconds across all three drives, NDRA engagement at TOR, reason for TOR.

Before deactivating the ADF after a TOR, the participant always looked straight ahead to check traffic in front of him, but he looked in the rear-view mirrors in only four cases. Consequently, in 9 out of 13 take-over scenarios, his checking and take-over behaviour did not meet the instructions that had been given before each drive and could – depending on the traffic situation – be considered as unsafe.

The participant disengaged the ADF at his own request seven times during the first drive and once during both, the second and third drive. The video analysis investigated the circumstances of these deactivations. During the first drive, the participant had his hands hovering over the steering wheel before three deactivations. In two cases, he overtook lorries after disengaging the ADF. One deactivation occurred before a construction site that otherwise would have triggered a system TOR. For the other driver-initiated take-over, no reason could be found in the video analysis. The deactivation in the second drive occurred before a motorway exit that otherwise would have triggered a system TOR. During the third drive, the participant disengaged the ADF after a driving error of the wizard (swerving in lane after a bump on the road surface), but activated the ADF soon afterwards again.

4.6.3 Participant C

Participant C was in the treatment group and experienced three TORs due to alleged system disturbances in the second drive.

At 97-98%, the participant's use of the ADF was very high in all of the three drives. While barely engaging in NDRAs in the first drive, a clear increase was observed in the second and third drive. As NDRAs, he chose to read magazines and use his smartwatch. While acceptance scores were high throughout all drives, trust was low in the beginning and increased over time. Take-over times ranged from 2.8 to 7.5 seconds and increased over time. Reading a magazine at the TOR appeared to prolong the take-over process. The three TORs due to an alleged system disturbance did not seem to influence trust, acceptance, or take-over times.

During the first drive, the wizard performed an emergency braking manoeuvre in automated mode after a lorry cut in closely in front of the vehicle.

During the third drive, the participant used reading glasses three times to read a magazine.

4.6.3.1 ADF use, NDRA engagement and gaze direction (RQ-U1, RQ-U9)

The participant's use of the ADF was very high in all three drives. While barely engaging in NDRAs during the first drive, a clear increase was observed in the second and third drive. When using NDRAs, he mostly looked at them too, but still carried out monitoring gazes. Participant C was the only one to use a smartwatch during ADF use.

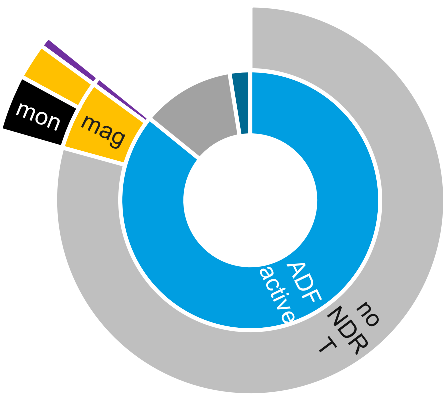

Figure 4.17: 1st drive of participant C

Participant C used the ADF 97% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged comprised five activations with durations between approx. 3 and 16.5 minutes (8 minutes on average). The participant spent about 7% of that time reading a magazine (5 sessions, yellow, middle layer) and less than 1% using his smartwatch (1 session, violet). While reading the magazine, he mainly monitored the ADF (black, outer layer), but also looked at his smartwatch while using it (violet) (overall 54.4% monitoring gazes). For the remaining time, he did not perform a NDRA.

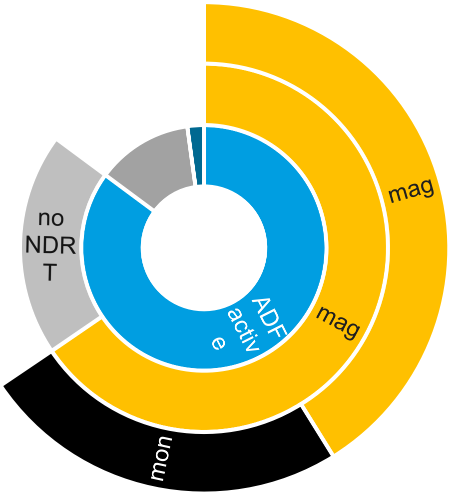

Figure 4.18: 2nd drive of participant C

Participant C used the ADF 97% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged comprised eight activations with durations between approx. 2.5 and 19 minutes (6 minutes on average). The participant spent about 77% of that time reading a magazine (9 sessions, yellow, middle layer). While reading the magazine, he mainly looked at the magazine (outer layer, 37.2% monitoring gazes). For the remaining time, he did not perform a NDRA.

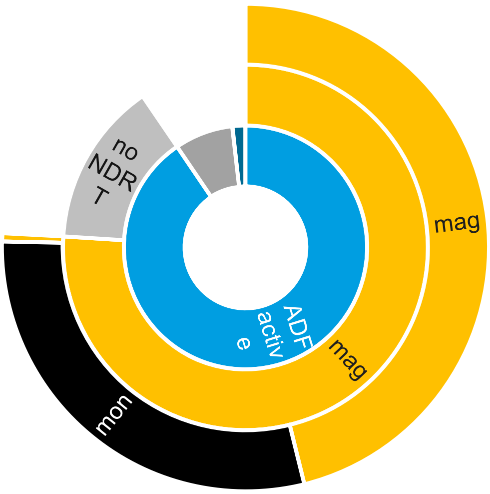

Figure 4.19: 3rd drive of participant C

Participant C used the ADF 98% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged comprised five activations with durations between approx. 3 and 18.5 minutes (9 minutes on average). The participant spent about 84% of that time reading a magazine (5 sessions, middle layer, yellow). While reading the magazine, he mainly looked at the magazine (outer layer, 38.5% monitoring gazes). For the remaining time, he did not perform a NDRA.

4.6.3.2 Trust and acceptance over time (RQ-U3, BASt-RQ 1, BASt- RQ 2)

The participant's self-reported trust was low before the first drive, but increased over time and reached high levels. Adaptation to the system can be assumed. After the second drive with three system disturbances, the trust level was the highest of all measurement points.

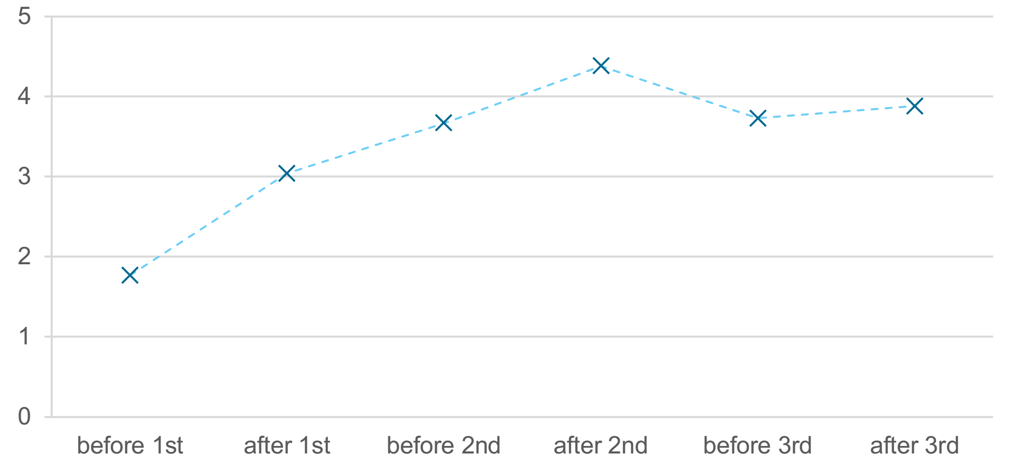

Figure 4.20: Development of trust over time (higher values indicate higher trust).

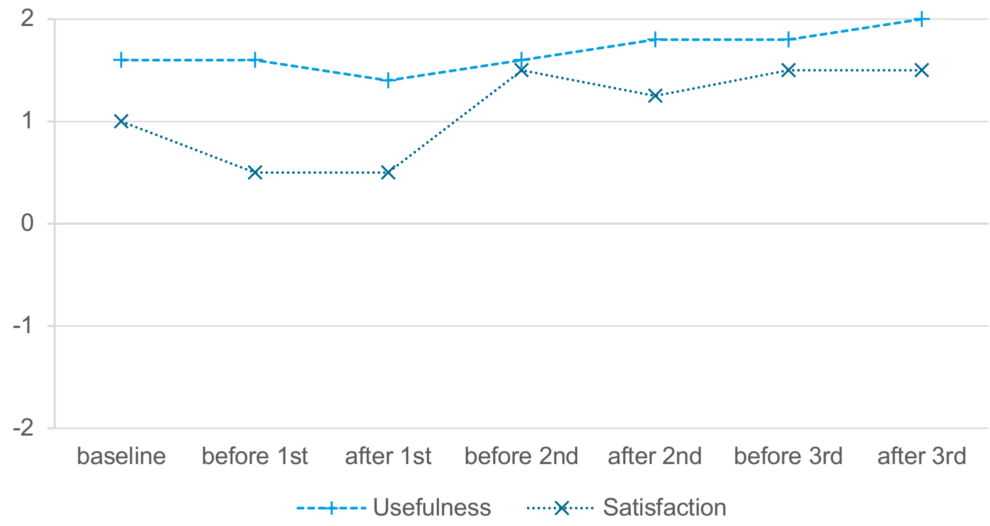

Figure 4.21: Acceptance rating scales (usefulness and satisfaction) over time (higher values indicate higher usefulness/satisfaction).

The levels of acceptance (usefulness and satisfaction scale) were high across all three drives, but especially satisfaction ratings increased before the second drive. There were only minor variations in acceptance after the second drive that had three system disturbances.

4.6.3.3 Take-over behaviour (RQ-U10, RQ-U11)

Across all three drives, the participant had to respond to 18 TORs. Over time, a tendency to longer take-over times was observed. In most (14 out of 18) take-over situations, the participant read a magazine. The three TORs due to a system disturbance did not differ from the other take-over scenarios.

Before disengaging the ADF, the participant always looked straight ahead to check traffic in front, butlooked in the rear-view mirrors in only eight cases. Consequently, in 10 out of 18 take-over scenarios, his checking behaviour did not meet the instructions that had been given before each drive and could – depending on the traffic situation – be considered unsafe.

There were no participant-initiated deactivations of the ADF during any drive.

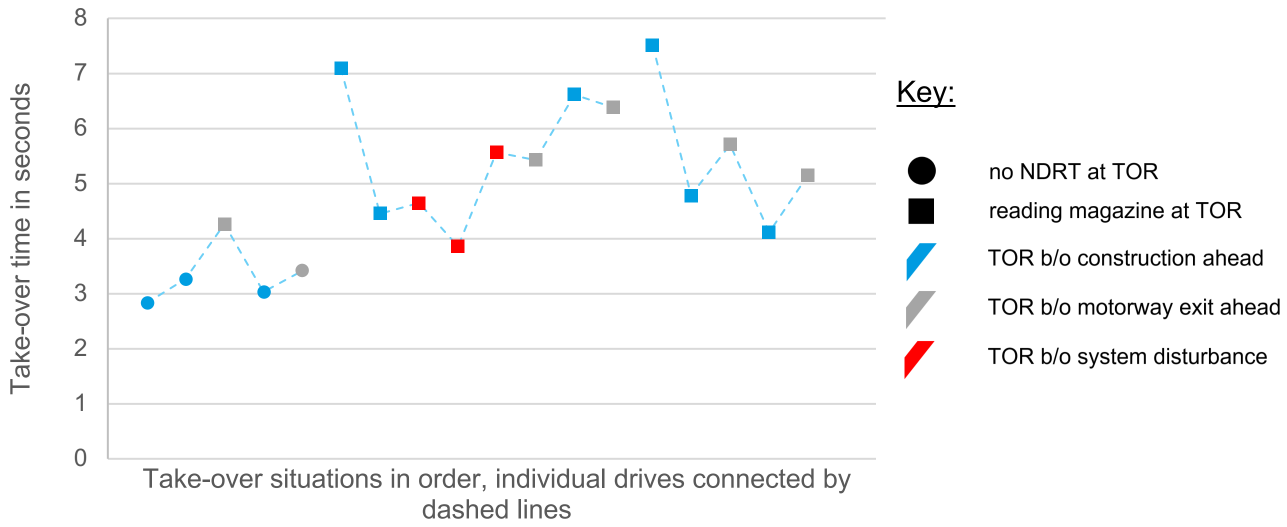

Figure 4.22: Take-over time in seconds across all three drives, NDRA engagement at TOR, reason for TOR.

4.6.4 Participant D

Participant D was initially assigned to the treatment group. Near the end of the first drive, the HMI system, designed to show the system status to the participant, froze and could not be used anymore. The drive was completed safely, but it was decided to transfer the participant to the control group and not let him experience other (alleged) system disturbances. As a result, participant D experienced a real, uncritical system disturbance during the first drive, but none of the manipulations planned for the treatment group.

The participant's use of the ADF was very high in all three drives (98-99%). In all three drives, he spent nearly all of the time in automated mode with NDRAs, namely tablet and smartphone use, reading magazines and engaging in office work. Both trust and acceptance scores were on a fairly high level from the beginning. Take-over times ranged from 3.8 to 7.8 seconds and showed a tendency to increase over time.

For the second drive, some of the video data could not be recorded. Therefore, some video analyses were limited.

4.6.4.1 ADF use, NDRA engagement and gaze direction (RQ-U1, RQ-U9)

The participant's use of the ADF was very high in all three drives. In all three drives, he spent nearly all of the time in automated mode with NDRAs, mostly looking at them, too. Participant D was the only one who performed office work as a NDRA (third drive): He had a notebook on one of his legs and his smartphone on the other and took notes of what he read on his phone.

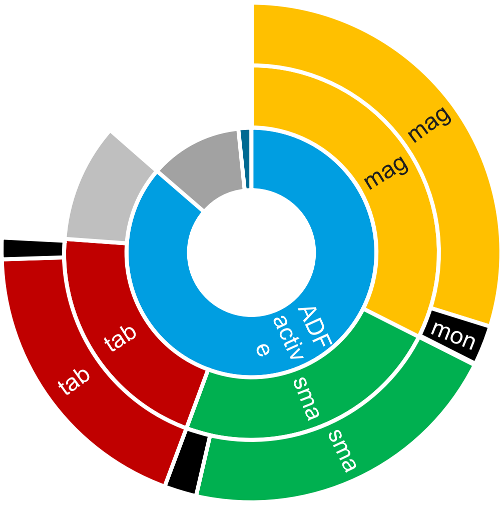

Figure 4.23: 1st drive of participant D

Participant D used the ADF 98% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged comprised five activations with durations between approx. 3.5 and 15 minutes (8.5 minutes on average). The participant spent about 37% of that time reading a magazine (2 sessions, yellow, middle layer), 27% using his smartphone (2 sessions, green) and 24% using a tablet (5 sessions, red). While engaging in the NDRAs, he mainly looked at them, too (outer layer). Monitoring gazes were relatively short (black, outer layer, overall 7.9%). For the remaining time, he did not perform a NDRA.

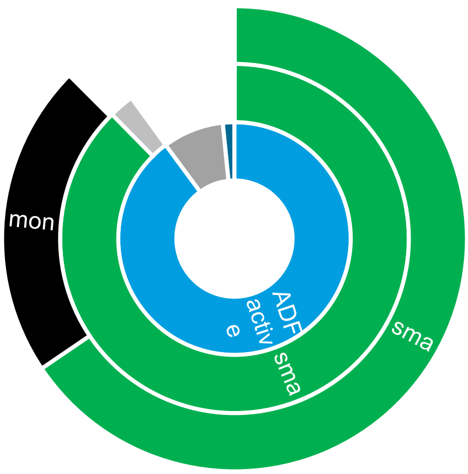

Figure 4.24: 2nd drive of participant D

Participant D used the ADF 98% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged comprised six activations with durations between approx. 1 and 18.5 minutes (7 minutes on average). The participant spent about 97% of that time using his smartphone (6 sessions, green, middle layer). While using his smartphone, he mainly looked at it, too. He monitored the ADF noticeably longer than in the first drive (24.9%). For the remaining time, he did not perform a NDRA.

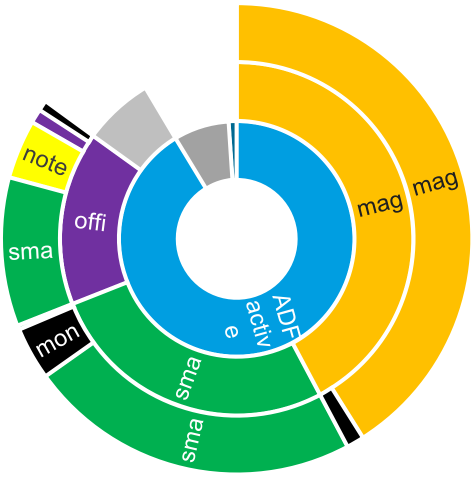

Figure 4.25: 3rd drive of participant D

Participant D used the ADF 99% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged comprised five activations with durations between approx. 3.5 and 18 minutes (9 minutes on average). The participant spent about 46% of that time reading a magazine (2 sessions, yellow, middle layer), 29% using a smartphone (5 sessions, green) and 17% performing office work (5 sessions, violet). During office work (using smartphone and notebook simultaneously), the participant mainly looked at his smartphone (green, outer layer). For all three NDRAs, the monitoring gazes add up to only a small number of gazes (6.6%). For the remaining time, he did not perform a NDRA.

4.6.4.2 Trust and acceptance over time (RQ-U3, BASt-RQ 1, BASt- RQ 2)

The participant's self-reported trust level was fairly high, barely varied and did not show a trend or signs of further adaptation to the system. After the first drive with the real and unplanned system disturbance, trust was only slightly lower than before.

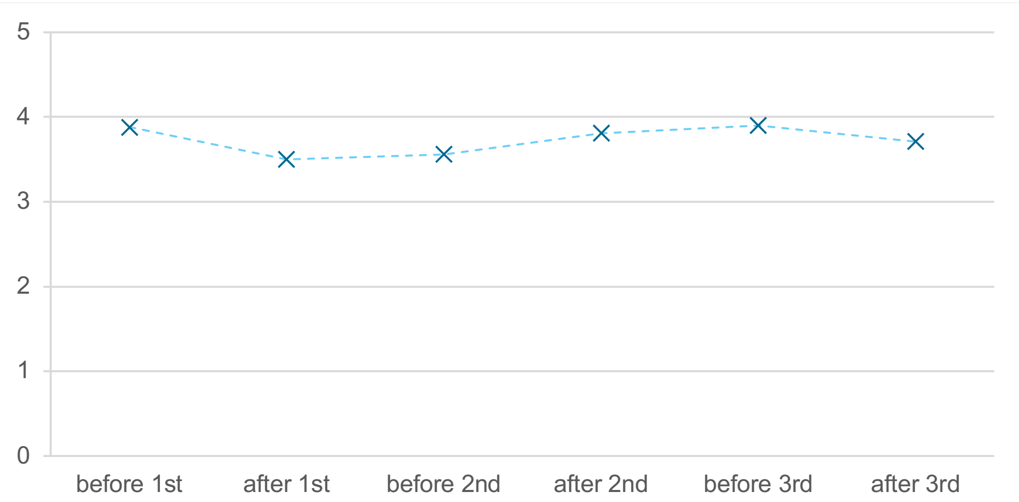

Figure 4.26: Development of trust over time (higher values indicate higher trust).

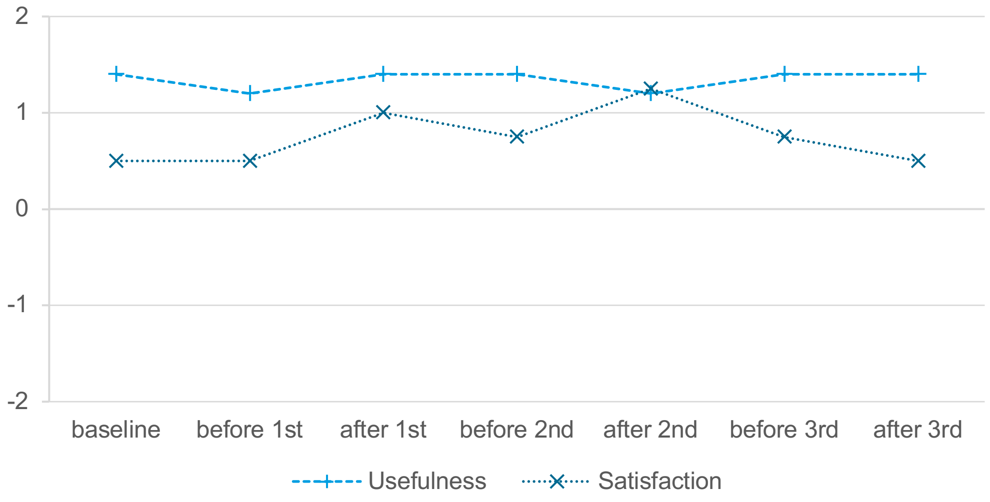

Figure 4.27: Acceptance rating scales (uesfulness and satisfaction) over time (higher values indicate higher usefulness/satisfaction).

The levels of acceptance (usefulness and satisfaction scale) were med-range to high across all three drives with only moderate variations. Satisfaction with the ADF ratings were mostly lower than usefulness ratings. After the first drive with the real and unplanned system disturbance, usefulness and satisfaction were rated slightly higher than before.

4.6.4.3 Take-over behaviour (RQ-U10, RQ-U11)

Across all three drives, the participant had to respond to 16 TORs. Over time, a tendency for longer take-over times was observed. In most (12 out of 16) take-over situations, the participant read a magazine.

Before disengaging the ADF, the participant always looked straight ahead to check traffic in front, but he looked in the rear-view mirrors in only ten cases. Consequently, in 6 out of 16 take-over scenarios, his checking behaviour did not meet the instructions that had been given before each drive and could – depending on the traffic situation – be considered unsafe.

There were no participant-initiated deactivations of the ADF during any drive.

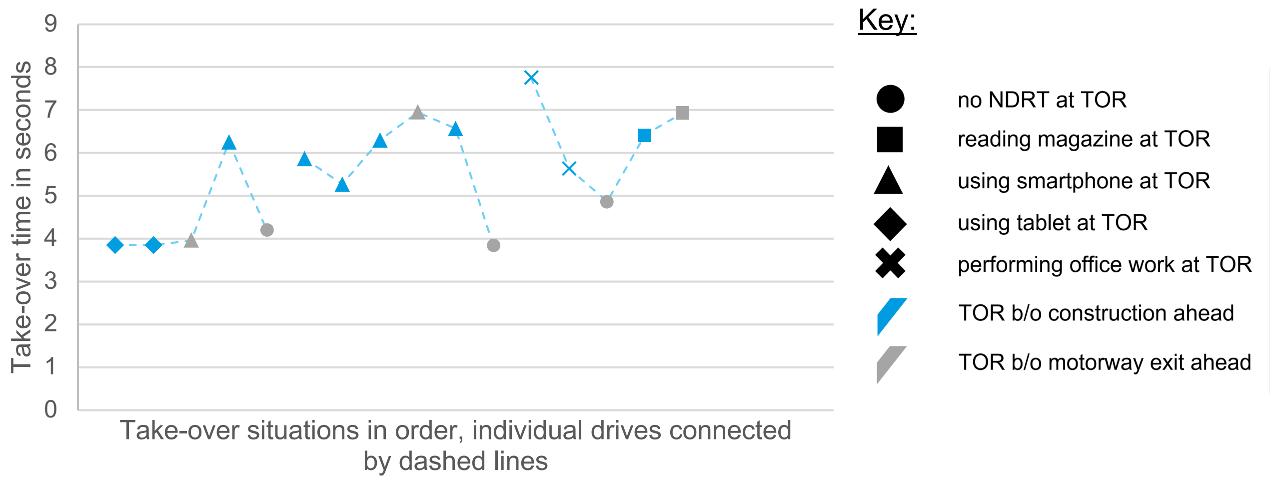

Figure 4.28: Take-over time in seconds over all three drives, NDRA engagement at TOR, reason for TOR.

4.6.5 Participant E

Participant E was in the control group and did not experience the alleged system disturbances.

The participant's use of the ADF was very high in all three drives (88-99%). Only during the second drive did he spent the majority of the time in automated mode with NDRAs. He engaged in smartphone and tablet use as well as reading magazines. Both trust and acceptance scores were medium to high over the three drives. Take-over times ranged between 1.6 and 3.2 seconds and did not show a clear trend. Reading a magazine at the TOR seemed to prolong the take-over process. In four situations, the participant chose to disengage the ADF.

The participant suspected that the car was controlled by a second human driver after the first drive but the examiner convinced him that it was a real ADF.

4.6.5.1 ADF use, NDRA engagement and gaze direction (RQ-U1, RQ-U9)

The participant's use of the ADF was very high in all three drives; however, he had the lowest duration of ADF engagement of all participants (88%) during his first drive. Only during the second drive did he spent more than half of the time in automated mode with NDRAs. In parts of the second drive, time spent with the magazine did not consist of actual reading; instead, the participant seemed to try to provoke reactions from other road users by holding the magazine up high to the side window or over the steering wheel. This obstructed the view of the wizard. Using a pretext, the examiner therefore asked the participant to take the magazine down.

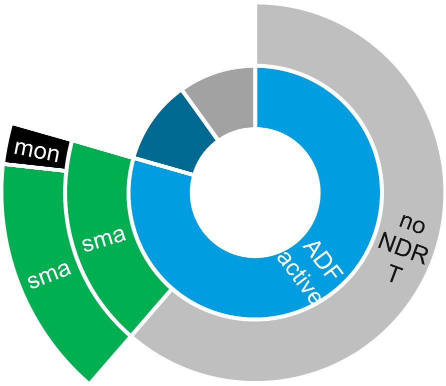

Figure 4.29: 1st drive of participant E

Participant E used the ADF 88% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged consisted of six activations with durations between approx. 2 and 15.5 minutes (6.5 minutes on average). The participant spent about 23% of that time using his smartphone (11 sessions, green, middle layer). During this time, he mainly looked at his smartphone (outer layer, 14.5% monitoring gazes). For the remaining time, he did not perform a NDRA.

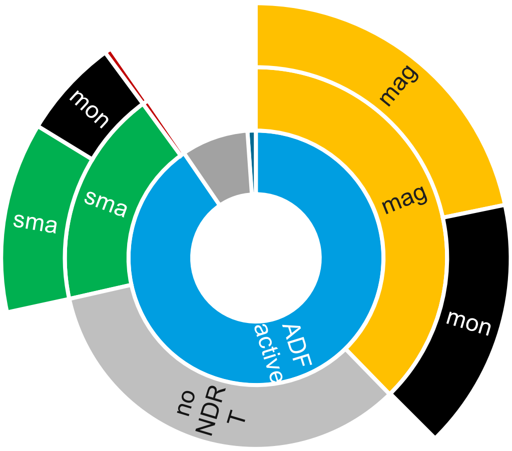

Figure 4.30: 2nd drive of participant E

Participant E used the ADF 99% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged consisted of five activations with durations between approx. 3 and 19 minutes (9 minutes on average). The participant spent about 42% of that time reading a magazine (10 sessions, yellow, middle layer), 20% with smartphone use (7 sessions, green) and less than 1% with tablet use (1 session, red). While engaging in NDRAs, he mainly looked at them (outer layer), but monitoring gazes (39.4%) took up more time than during the first drive. For the remaining time, he did not perform a NDRA.

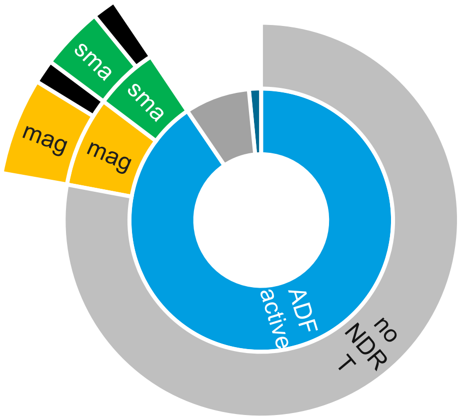

Figure 4.31: 3rd drive of participant E

Participant E used the ADF 98% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged consisted of six activations with durations between approx. 26 seconds and 19.5 minutes (7.5 minutes on average). The participant spent about 8% of that time reading a magazine (3 sessions, yellow, middle layer) and 6% with smartphone use (1 session, green). While engaging in the NDRAs, he mainly looked at them, too (outer layer). Monitoring gazes were relatively short (black, outer layer, 22.6%). For the remaining time, he did not perform a NDRA.

4.6.5.2 Trust and acceptance over time (RQ-U3, BASt-RQ 1, BASt- RQ 2)

The participant's self-reported trust level was fairly high, varied moderately and did not show a clear trend or signs of further adaptation to the system.

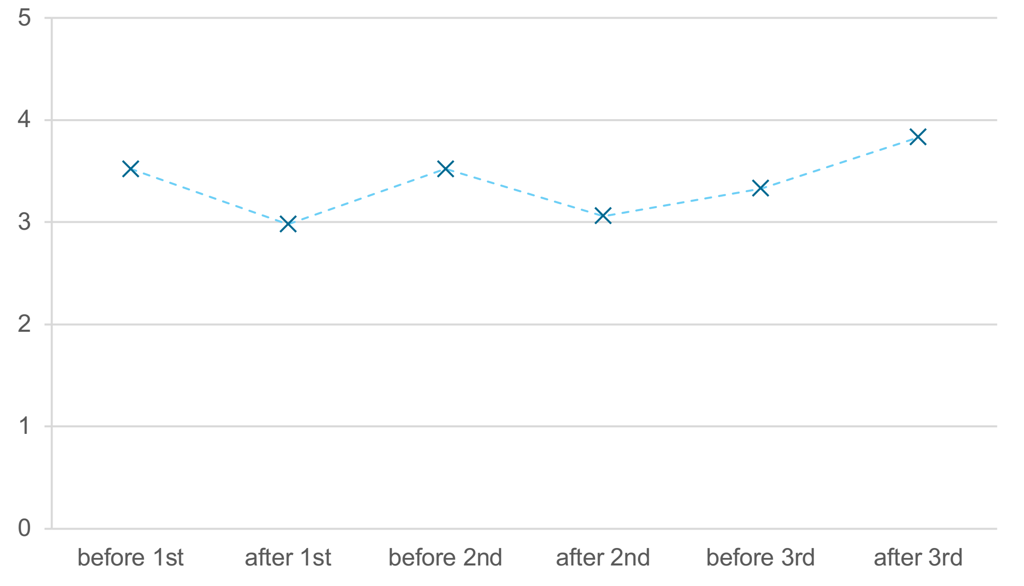

Figure 4.32: Development of trust over time (higher values indicate higher trust).

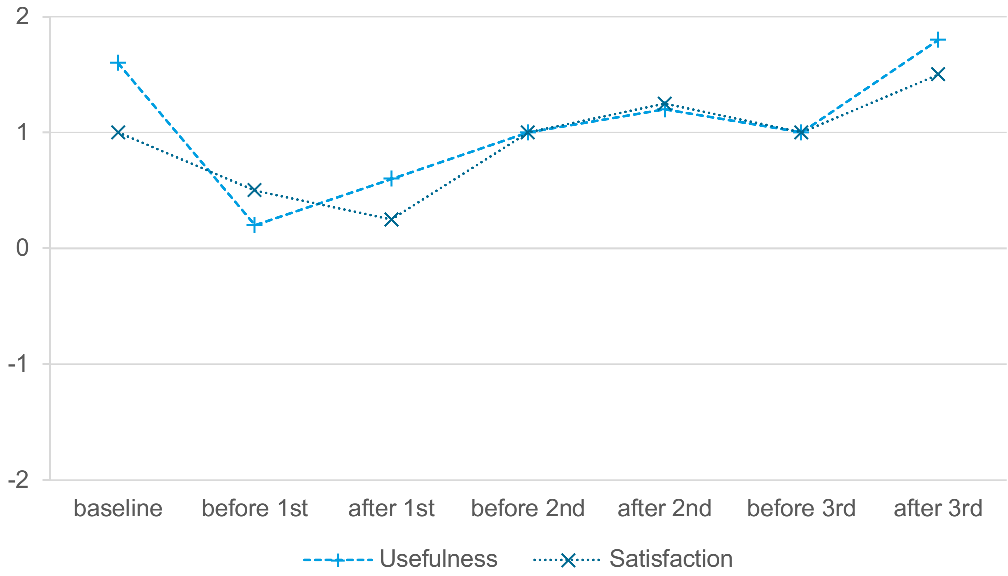

Figure 4.33: Acceptance rating scales (usefulness and satisfaction) over time (higher values indicate higher usefulness/satisfaction).

The levels of acceptance (usefulness and satisfaction scale) were medium to high over all three drives. During the first drive, acceptance was lowest andshowed an upwards trend after that.

4.6.5.3 Take-over behaviour (RQ-U10, RQ-U11)

Across all three drives, the participant had to respond to 13 TORs in total. There were no apparent training effects with regard to take-over time. When the participant was engaged in reading a magazine, his take-over time was among the highest of all of his take-over scenarios.

Before disengaging the ADF, the participant always looked straight ahead to check traffic in front, but he never looked in the rear-view mirrors. Consequently, in all of his 13 take-over scenarios, his checking and take-over behaviour did not meet the instructions that had been given before each drive and can – depending on the traffic situation – be considered unsafe.

The participant chose to disengage the ADF three times during the first drive and once during the third drive. The video analysis investigated the circumstances of these deactivations: In the first drive, two deactivations occurred before a construction site or motorway exit that otherwise would have triggered a system TOR. For the remaining deactivation in the first drive and the one in the third drive, the reasons were not obvious.

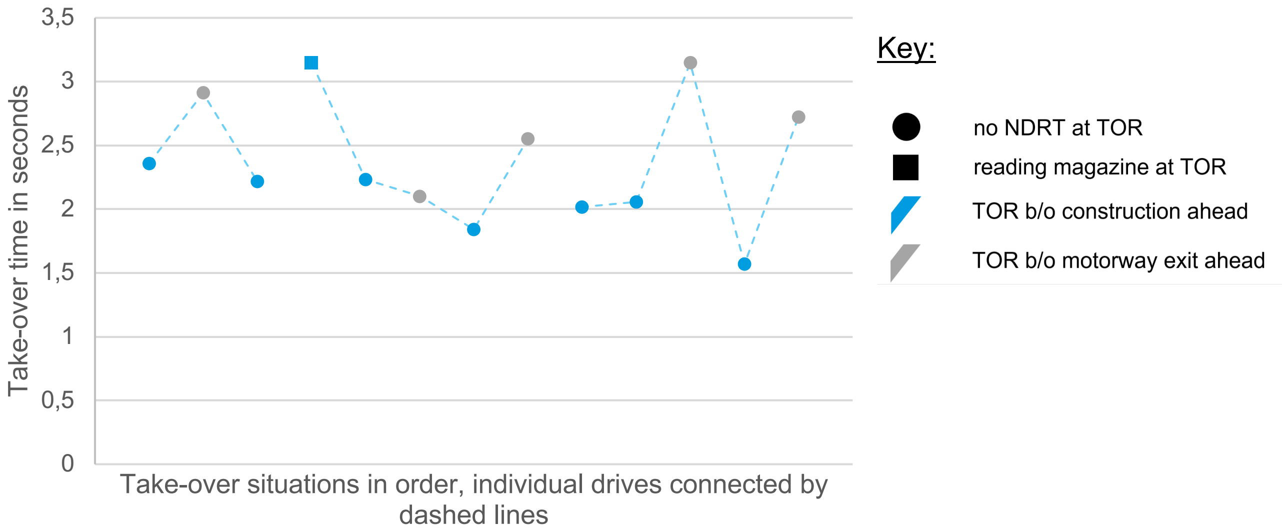

Figure 4.34: Take-over time in seconds over all three drives, NDRA engagement at TOR, reason for TOR.

4.6.6 Participant F

Participant F was in the control group and did not experience any system disturbances.

The participant's use of the ADF was very high in all three drives (98-99%). In all three drives, he spent the vast majority of the time in automated mode with NDRAs. As NDRAs, he chose magazines and smartphone use. Both trust and acceptance scores were medium to high across the three drives. Take-over times ranged from 2.4 to 5.3 seconds and did not show a clear trend over time.

4.6.6.1 ADF use, NDRA engagement and gaze direction (RQ-U1, RQ-U9)

The participant's use of the ADF was very high in all three drives. In all three drives, he spent nearly all of the time in automated mode reading a magazine.

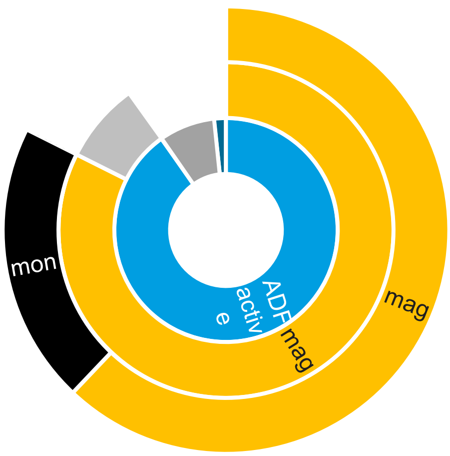

Figure 4.35: 1st drive of participant F

Participant F used the ADF 98% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged consisted of five activations with durations between approx. 1.5 and 18 minutes (8 minutes on average). The participant spent about 91% of that time reading a magazine (4 sessions, yellow, middle layer). During this time, he mainly looked at it, too (outer layer, 24.7% monitoring gazes). For the remaining time, he did not perform a NDRA.

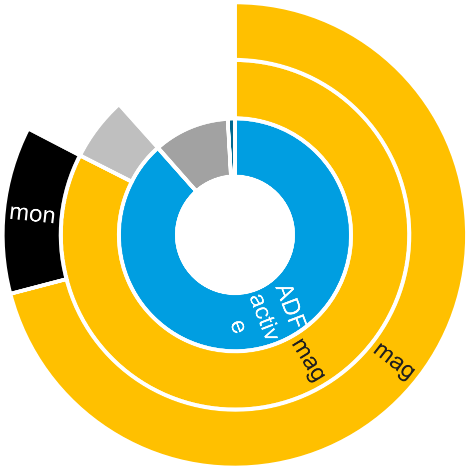

Figure 4.36: 2nd drive of participant F

Participant F used the ADF 99% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged consisted of five activations with durations between approx. 3.5 and 15 minutes (8.5 minutes on average). The participant spent about 93% of that time reading a magazine (5 sessions, yellow, middle layer). During this time, he mainly looked at it, too (outer layer). Monitoring gazes (14.0%) were reduced compared to the first drive. For the remaining time, he did not perform a NDRA.

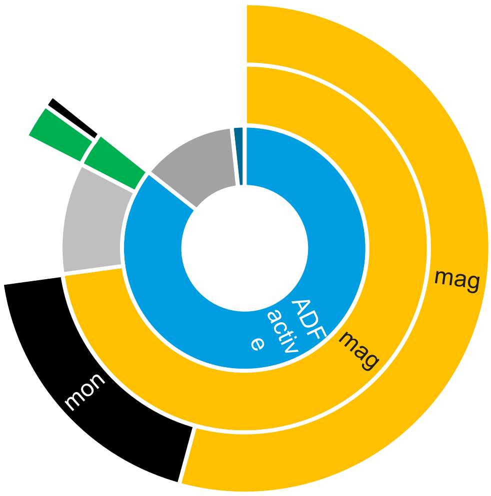

Figure 4.37: 3rd drive of participant F

Participant F used the ADF 98% of the time (lighter blue, inner layer) that it was available (darker blue). The time with ADF engaged consisted of six activations with durations between approx. 3 and 18 minutes (7 minutes on average). The participant spent about 85% of that time reading a magazine (5 sessions, yellow, middle layer) and under 4% with smartphone use (2 sessions, green). During this time, he mainly looked at the respective NDRAs (outer layer, 25.3% monitoring gazes). For the remaining time, he did not perform a NDRA.

4.6.6.2 Trust and acceptance over time (RQ-U3, BASt-RQ 1, BASt- RQ 2)

The participant's self-reported trust level was fairly high, nearly constant and did not show a trend or signs of further adaptation to the system.

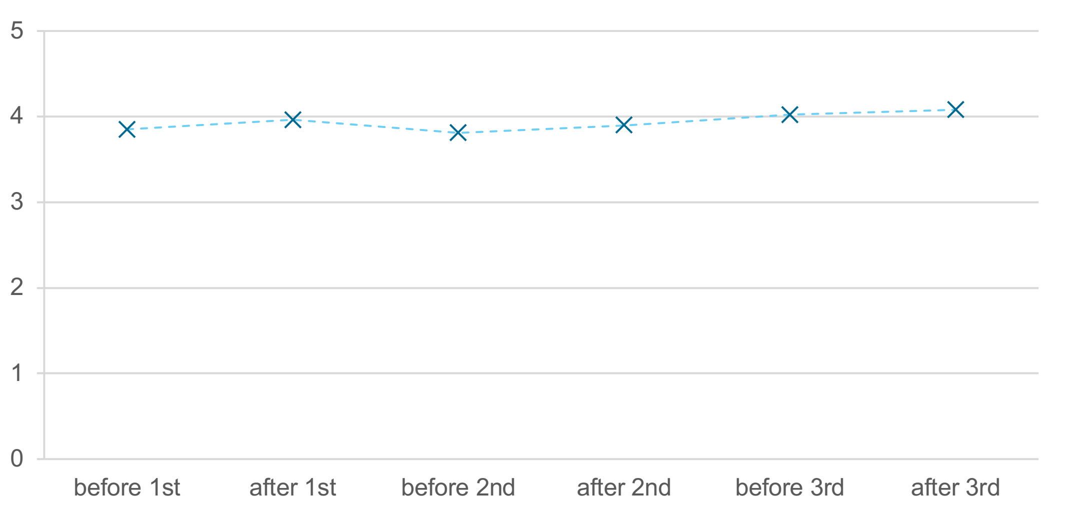

Figure 4.38: Development of trust over time (higher values indicate higher trust).

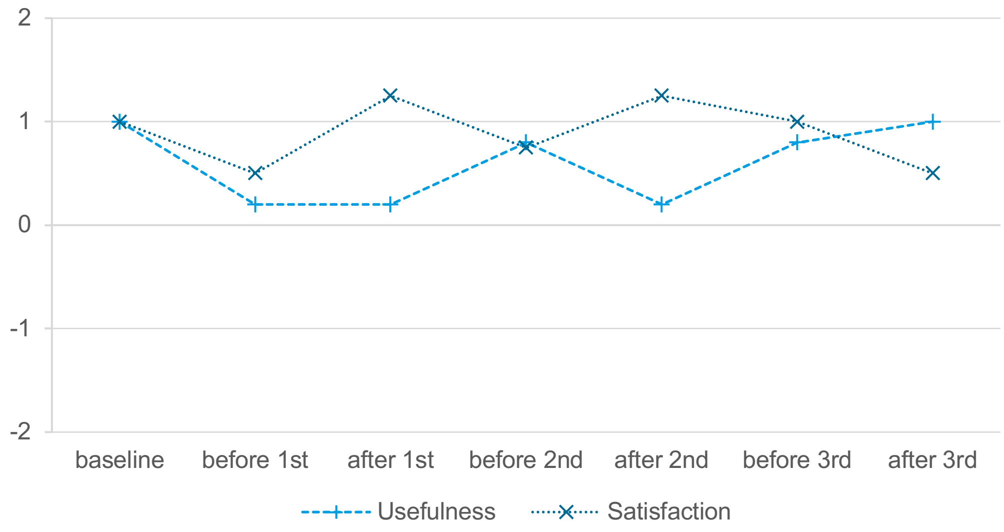

Figure 4.39: Acceptance rating scales (usefulness and satisfaction) over time (higher values indicate higher usefulness/satisfaction).

The levels of acceptance (usefulness and satisfaction scale) were medium to high over all three drives. They varied over time, but showed no clear trend.

4.6.6.3 Take-over behaviour (RQ-U10, RQ-U11)

Across all three drives, the participant had to respond to 16 TORs. There were no apparent training effects with regard to take-over time. When the participant was engaged in reading a magazine, it took him more time to take over than in scenarios without NDRA.

Before disengaging the ADF, the participant always looked straight ahead to check traffic in front of him, but he looked in the rear-view mirror in only seven cases. Consequently, in 9 out of 16 take-over scenarios, his checking and take-over behaviour did not meet the instructions that had been given before each drive and could – depending on the traffic situation – be considered unsafe.

There were no participant-initiated deactivations of the ADF during any drive.

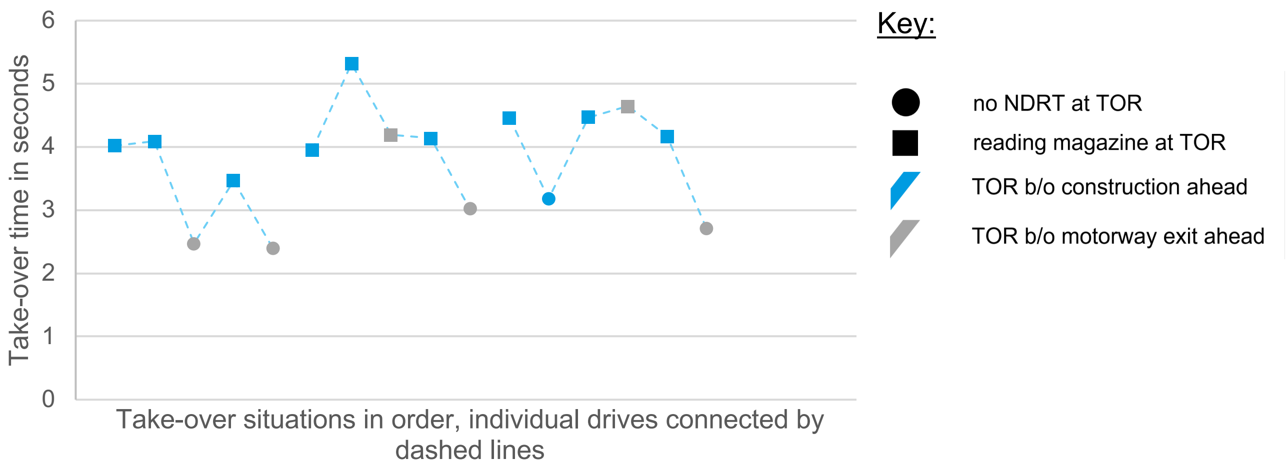

Figure 4.40: Take-over time in seconds over all three drives, NDRA engagement at TOR, reason for TOR.

4.7 Comparison of Results

The data, behaviour and questionnaire data recorded during this Wizard of Oz study with six participants showed similarities and differences.

While being free to switch the ADF on and off as they pleased, participants activated the ADF in 88-99% of the time that it was available. The number of ADF activations per drive varied between 5 and 11 and the duration of these activations ranged between 13 seconds and approximately 19.5 minutes. Different traffic situations and driver-initiated take-overs influenced both the number and duration of activations.

For three of the six participants (participants A, B, C), an increase in NDRA engagement over time was observed. Two participants (D, F) spent the vast majority of automated driving time in all three drives on NDRAs, so that there was barely room for a further increase. Participant E spent very little time on NDRAs during his first and third drive, but more than half of the time during the second drive. The choice of NDRAs varied widely between and also within participants: Smartphone use and magazines were generally popular (participants A, D, E, F used both of these at some time; B and C used their smartphone, but not magazines). The tablet was used by two participants (D, E); smartwatch use (C), drinking (B) and office work (D) were observed less often.

Monitoring gazes towards the road, the rear-view mirror or the instrument cluster were observed for all participants during NDRA engagement, but durations and time portions differed. Monitoring gazes were longer than gazes towards the respective NDRA only during the first drive of participant C and while participant B was drinking during his second and third drive. Of all participants, Participant C monitored the ADF the most (38.5%), participant D the least (13.4%) (both relative to the time spent with NDRAs over all three drives). A clear trend for time spent on monitoring gazes was only visible for participant C (decreasing from 54.4% to 38.5%).

For all but participant C, trust in the ADF was always fairly high or high. Participant C developed high trust over time and reached the trust level of the other participants before the second drive. Neither the planned system disturbances of the treatment group (A, C) nor the real and unplanned system disturbance (D) had an obvious influence on trust.

Acceptance towards the ADF attracted medium ratings at least, but usually high ratings at all measurement points. Neither the planned system disturbances of the treatment group (A, C) nor the real and unplanned system disturbance (D) had an obvious influence on acceptance.

When being prompted by the system to take over, all participants were able to do so in under 10 seconds and thus, within the provided time budget. The take-over times differed widely both between and within the participants, ranging from 0.78 seconds (participant B, third drive) to 7.76 seconds (participant D, third drive). For participant B, training effects with regard to take-over times can be assumed because of decreasing take-over times. The opposite trend was found for C and D, where take-over times tended to increase over time. In this case, the adaptation to the system could have caused the increase, similar to a training effect. For participants A, E and F, take-over times appeared erratic or stayed within a certain range. Reading a magazine at the TOR seemed to prolong take-over times; this was the case for all participants to a certain extent.

None of the participants checked all of their take-overs in accordance with instructions: the look in the rear-view mirror was often missing.

Two participants chose to deactivate the system. Reasons for this varied, e.g. participant B overtook lorries after the deactivation ; participants B and E deactivated the ADF shortly before road works or a motorway exit would have triggered a system-initiated TOR.

4.8 Discussion

The results of this study paint a positive picture of users' trust and acceptance in automated vehicles overall. The high usage of the ADF and overall high trust scores correspond with mostly high acceptance ratings. Besides the possible influence of the examiner's and second driver's presence on trust, it seems that the participants indeed trusted the ADF and experienced it as satisfying and useful.

Trust and acceptance of two participants in the treatment group seemed to be unaffected by the three non-critical system disturbances. Generalisation is not advisable due to the group size, but the findings can serve as a basis for deeper investigation of system disturbances which cannot be completely ruled out in in-production vehicles with SAE Level 3 ADFs.

Partipcipants' NDRA use varied considerably. Smartphone use and reading magazines seemed to be particularly popular. The former was to be expected because of the breadth of functionalities smartphones provide, ranging from communication to entertainment or information. Since all magazines read by the participants were provided by BASt, it is uncertain if drivers in mass-produced automated vehicles would also read magazines during automated driving. Perhaps, they would engage more in tasks more easily available to them, e.g. via smartphone. Beyond smartphones, one participant brought his own NDRA (office work). It should be investigated how drivers would chose to spend time in automated mode after several weeks or months of use to explore if and how their chosen NDRAs changed over time.

The study revealed possible problems of user's behaviour and interaction with automated vehicles: Checking traffic behind through the rear-view mirrors in a take-over situation often did not match the instructions that had been given before each drive. Depending on the traffic situations, this can compromise traffic safety. Future research should investigate how drivers' checking behaviour can be improved and maintained at a high standard even after months or years of use. One participant showed signs of microsleep during automated driving which could be a sign of overtrust. Technical systems such as driver-monitoring-systems and education on possible risks associated with driver states in automated vehicles should be developed in order to mitigate the risk posed by tired drivers. One participant was distracted by his smartphone while driving manually. The possibility to engage in NDRAs legally and safely during automated driving could entice drivers to continue their use during manual driving. Driver-monitoring-systems and driver education could be instruments to prevent drivers from continuing to use an NDRA in manual driving mode. Another participant tried to provoke reactions from other road users by holding a magazine up high to the side window of the vehicle. Especially incoming years when Level 3 automated vehicles are sold but have not yet reached high market penetration and are therefore a unfamiliar to most road users information campaigns on the use of NDRAs could help prevent misunderstandings between users of automated vehicles and conventional car users.

4.9 Limitations

The utilisation of a Wizard of Oz vehicle has already been proven to work in previous studies on human machine interaction with automated vehicles. In absence of a real ADF this reliable and flexiable technique gives the participants a realistic and credible experience of automated driving on public roads. However, the presence of both the examiner and the second driver during the drives could have influenced the participants' trust: The alleged ADF could have been perceived as generally safe, since two employees of BASt were also in the vehicle. Future studies should investigate trust in automated vehicles without researchers or engineers on board.

This study was designed as a long-term study of the development of trust and acceptance in users. Therefore, the participants experienced the ADF three times with intervals of one week between each drive. The approach was successful, as changes in behaviour (e. g. NDRA engagement, take-over times, rise of trust levels) were clearly visible in participants. Studies that investigate trust and acceptance over months of use could further extend our knowledge.

The participants came from a demographically homogenous group and the sample size was small. Despite these limitations, observed behaviour varied considerably. Nonetheless, trust and acceptance of other age groups and genders should be investigated for the holistic understanding of users' perception of automated driving. An online study, conducted by BASt aimed to gain deeper insights in users' acceptance and NDRA engagement in automated vehicles (see chapter 11).

5 Driving simulator study on long-term behavioural adaptation

In this chapter a driving simulator study is described that explored changes in acceptance and usage of an L3/L4-ADF with growing experience. The work was conducted by WIVW.

5.1 Aim and research questions

The main focus of this simulator study was the investigation of behavioural adaptation (BA) with repeated usage of a motorway ADF. Drivers in the study experienced an L3/L4-ADF six times in the driving simulator. Driver-related concepts like system usage, acceptance or trust were assessed and their change over time was analysed. The analysis of BA was done for following L3Pilot RQs:

- RQ-U1: Are drivers willing to use an ADF?
- RQ-U3: What is the user' acceptance of the ADF?
- RQ-U4: What are drivers' expectations regarding system features?
- RQ-U5: What is the impact of ADF on driver state?
- RQ-U6: What is the impact of ADF use on driver awareness?
- RQ-U9: What is drivers' secondary task engagement during ADF use?
- RQ-U10: How do drivers respond when they are required to retake control?
- RQ-U11: How often and under which circumstances do drivers choose to activate/deactivate the ADF?

Furthermore, the simulator study planned to investigate several RQs that were not part of the common list of RQs from D3.1. One specific focus was the special use case of a drowsy driver, i.e. the question how acceptance and usage of an ADF is influenced by driver fatigue. The other was the impact of the automation level or system capability on the evaluation and usage of the system by drivers. In the experiment acceptance and usage of a Level 3 versus a Level 4 automated system implementation according to the SAE (2021) standard were compared.

5.2 Methods

5.2.1 Experimental approach

The basic idea of the study was that drivers should experience a number of drives with a highly automated motorway system in the driving simulator. Insights into changes over time in terms of drivers' behaviours and attitudes were expected. Various environmental conditions during system usage were implemented in terms of traffic density, infrastructure, and weather.

Also, the system implementation varied with regard to the automation level. Most of the automated motorway systems tested in the pilots were defined as automation L3 systems. However, L4 systems were also included in the overall scope of L3Pilot and in some of the on-road tests. It is reasonable to expect their introduction o the roads in a not-so-distant future. It is therefore of interest to investigate drivers' acceptance and usage of AD systems depending on the level of automation.

One special issue is driver drowsiness, and fatigue was experimentally induced to study this: one experimental drive took place at 6 a.m. after a night of partial sleep deprivation. This setup has been proven to produce high levels of drowsiness and was highly sleep-inducing when driving with a L3/L4-system in previous studies (e.g. Wörle, Metz, Ottersen & Baumann, 2020; Wörle, Metz, Thiele & Weller, 2019).

To study user-related topics such as acceptance or willingness to use, the questionnaire developed within L3Pilot (see Metz et al., 2020) was used. Before and after every experimental drive, the drivers' subjective evaluation of the system was assessed with this questionnaire. The survey software LimeSurvey was used for this purpose.

5.2.2 Study environment

The study took place in a driving simulator with a motion system at WIVW. The simulator consists of a hexapod motion platform and is equipped with a mock-up consisting of the front half of a BMW fitted with original parts. It offers a surround view of 240 degrees, as well as displays that serve as left, right, and rear-view mirrors. The driving simulator runs with the simulation software SILAB®.

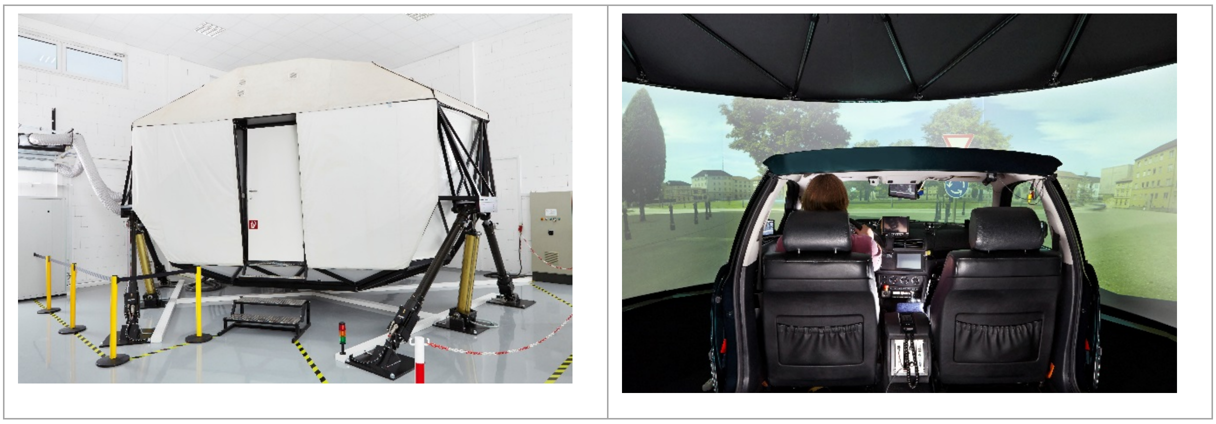

Figure 5.1: WIVW's motion-based driving simulator from the outside (left) and from the inside (right).

The data logging included signals from the driving simulator software that covered the areas of vehicle dynamics (v, ax, ay), the state of the L3/L4 system (TORs, system status), vehicle handling (brake pedal position, steering angle, hands-on detection), vehicle environment (distance to other vehicles, lane position), as well as continuous video recording of the driver and the driving scenery. The simulator was equipped with the four-camera remote eye-tracking system SmartEye Pro®. This system automatically recorded head position and movement, together with gaze direction and eyelid opening. Furthermore, the experimenter continuously coded whether the driver was engaging in non-driving related activities. The coding was done via a tablet application. The coding on the tablet was saved in synchronization with the rest of the data in one data log file.

5.2.3 Tested functions

The systems used in the driving simulator study were motorway functions with a speed range of 0 to 130 km/h. The systems adapted speed to surrounding traffic as well as to speed limits along the road. On sections with no speed limit, the system maintained a speed of 130 km/h. The system was able to execute lane changes automatically in order to overtake slower vehicles.

Two automation levels were implemented:

- Level 3: The system was implemented as an L3 automated motorway pilot. Drivers were instructed that they would not have to pay attention to the driving task in automated mode and could engage in other activities. However, when the system issued a TOR, they had to retake control of the vehicle and were responsible for the driving task. TORs were issued fairly frequently. All take-overs were issued with a time budget of 15 seconds. Although not mandatory for L3 functions, a minimal risk manoeuvre (MMR) was performed in the event that the driver did not take control back during the take-over time. In that case, the vehicle stopped in its lane.
- Level 4: The system was implemented as a L4 motorway chauffeur. Drivers were instructed that while in automated mode they would not have to pay attention to the driving task and could engage in other activities, because all driving situations could be handled by the system. TORs were issued with a large time budget of 45 seconds. If the driver failed to take over, the vehicle executed the MRM.

The implemented ODD for both systems was based on the definitions for the market-ready L3Pilot motorway pilot. This means that the ADF was not available in the following conditions:

- On exits from and entrances to motorways
- Through construction sites
- On longer sections with poor/missing lane markings
- In heavy rain

If the system approached one of the system boundaries, a TOR was issued to the driver. The timing of the TOR depended on the system (L3 v. L4). Missing/poor lane markings were considered to be outside ODD only for the L3 system; with L4 no TOR occurred. In the event that the driver did not respond to a TOR, an MRM was executed.

Both system variants were implemented without any unplanned failures. This means that all TORs were due to the defined system boundaries and the system was working fine within the boundaries of ODD.

5.2.4 Test drives

Six experimental drives were implemented in the driving simulator. They varied not only in length but also with regard to the type and duration of traffic situations experienced while driving in ODD as well as in the number of and reasons for TORs. The aim was to implement drives with a reasonable length (about 30 minutes) that contained everyday driving situations. Reasons for TORs were everyday situations on motorways such as highway intersections, construction sites or sections with bad lane markings. Unusual or critical situations were not included. Furthermore, two drives were specifically implemented to address fatigue. These drives were longer and comprised sections of monotonous highway driving with little traffic. Table 5.1 provides a summary of the content of the six test drives, including the number of and reasons for the TORs.

Table 5.1: Content of the test drives.

Drive A – 30 min			
Driving in ODD	Driving outside ODD	TOR L3	TOR L4
Section with low traffic density & changing speed limit Traffic jam	In parking area At motorway junction	1 x before highway intersection 1 x before exit 1 x poor lane markings	1 x before highway intersection 1 x before exit
Drive B – 30 min			
Driving in ODD	Driving outside ODD	TOR L3	TOR L4
Section with low traffic density & changing speed limit Traffic jam	In parking area In construction site	1 x construction site 1 x exit 2 x poor lane markings	1 x construction site 1 x exit
Drive C – 30 min			
Driving in ODD	Driving outside ODD	TOR L3	TOR L4
Section with low traffic density & changing speed limit Traffic jam	In parking area	1 x exit 1 x roadworks	1 x exit 1 x moving roadworks
Drive D – 30 min			
Driving in ODD	Driving outside ODD	TOR L3	TOR L4
Section with low traffic density & changing speed limit	In parking area At motorway junction In construction site	1 x construction site 1 x highway junction 1 x before exit 2 x poor lane markings	1 x construction site 1 x highway junction 1 x before exit

Drive E – 120 min			
Driving in ODD	Driving outside ODD	TOR L3	TOR L4
Section with low traffic density & changing speed limit	In parking area In heavy rain	1 x exit 1 x roadworks 1 x heavy rain	1 x exit 1 x roadworks 1 x heavy rain
Drive F – 120 min (with sleep deprivation)			
Driving in ODD	Driving outside ODD	TOR L3	TOR L4
Section with low traffic density & changing speed limit	In parking area In heavy rain	1 x exit 1 x roadworks 1 x heavy rain	1 x exit 1 x roadworks 1 x heavy rain

Drivers got to know the ADF in a short introduction drive, during which it was explained how to activate and deactivate the system and where drivers experienced the basic system behaviour (lane keeping, lane change and overtaking, TOR). The introducton drive started on an empty highway; drivers were shown how to turn the system on and off. Subsequently, they encountered a slower vehicle ahead in their lane, and drivers experienced a fully automated lane change. After that, there were two TORs without external reasons. Drivers were instructed not to react to the first TOR so that they could experience the system behaviour, including emergency stop. At the second TOR drivers were instructed to take back control and deactivate the function.

5.2.5 Experimental procedure

Drivers were invited to participate in a study on the long-term effects of an L3/L4 motorway chauffeur. In the introducton session drivers received information about the schedule for their test drives. Before every session, they knew the duration of the oncoming trip, and they were informed that they were free to prepare for the drive as they wished. This meant for instance that they could bring something to read, something to eat, or prepare other potential side tasks to fill the time of the automated drive. Table 5.2 gives an overview of the six experimental sessions. The order of the drives was varied in order to avoid sequence effects.

Table 5.2: Overview of the content of the six sessions of the experiment.

Session	Content
Session 1 – introduction session 60 min	• Information on experiment & planned schedule • Informed consent • Handing out L3Pilot questionnaire part 1 • Introduction drive (10 min) • Drive A or D (30 min) • Post-drive questionnaire (full version)

Session	Content
Session 2 45 min	• Short pre-drive questionnaire • Drive B or Drive C (30 min) • Post drive questionnaire (short version)
Session 3 150 min	• Short pre-drive questionnaire • Drive E (90 min) or Drive F (90 min) with sleep deprivation • Post-drive questionnaire (short version)
Sessions 4 45 min	• Short pre-drive questionnaire • Drive A or Drive B (30 min) • Post drive questionnaire (short version)
Session 5 150 min	• Short pre-drive questionnaire • Drive E (90 min) or Drive F (90 min) with sleep deprivation • Post-drive questionnaire (short version)
Session 6 90 min	• Short pre-drive questionnaire • Drive C or Drive D (30 min) • Post drive questionnaire (full version)

5.2.6 Study design

A sample of N = 61 drivers participated in total. Two experimental factors were varied in a between-subjects design (see Table 5.3):

- System implementation (L3 v. L4)
- Order of the drives (1 v. 2)

Table 5.3: Experimental design of the driving simulator study.

Experimental design: between design N = 60			
L3		L4	
Group 1 (N = 16)	Group 2 (N = 15)	Group 3 (N = 15)	Group 4 (N = 15)
Drive A	Drive D	Drive A	Drive D
Drive C	Drive B	Drive C	Drive B
Drive F	Drive E	Drive F	Drive E
Drive B	Drive A	Drive B	Drive A
Drive E	Drive F	Drive E	Drive F
Drive D	Drive C	Drive D	Drive C

5.3 Data Sources and Analysis

Based on the questionnaires, the drivers' evaluation of the system (concepts such as acceptance, trust, comfort, etc.) were analysed. In addition to the general L3Pilot questionnaire used in full version after the 1st and after 6th session, 20 items were presented that asked for the drivers' system understanding or mental model of the ADF. The items were statements about the ADF, and participants had to chose whether the statement was correct (Yes/No) or whether the did not know. The item list for the system understandning can be found in the annex.

Information collected via questionnaires and data logged during the drives were used to answer the research questions. Table 5.4 lists the different data sources logged during all experimental drives, together with the indicators that were derived from them.

Table 5.4: Data sources logged during the drives in the driving simulator and indicators derived from them.

Data source	Calculated indicators
Time series data logged from the driving simulator software SILAB®: • System status (on/off, TOR, availability) • Vehicle handlig (hands on) • Coding of driving situation (e.g. current lane, speed limit)	Proportion of time driving with system active overall and separately for difference scenarios • Driving in traffic jam • Free flow conditions with speed limit • Free flow conditions without speed limit • Stable driving on preferred lane • Driving on left lane (overtaking) Reaction times after TOR • Hands on time • Time until system is turned off
Video of driving scenery and driver's face	Evaluation of take-over performance via TOC rating
Continuous coding of secondary task engagement done by observer	Proportion of time spent on NDRAs Proportion of time spent on NDRAs involving both hands
Head position and movement and gaze direction measured with SmartEye Pro®	Proportion of glances directed to the road (PRC) Eyes on road time after TOR
Eye-lid opening level measured with SmartEye Pro®	Evaluation of driver fatigue based on proportions of time the eye is closed (PERCLOS)
EEG recording logged during monotonous drives	Proportion time spent sleeping (N1 + N2 + micro sleep)

5.4 Sample description

A total of N = 61 drivers participated in the simulator study of long-term effects on user acceptance. N = 31 drivers were assigned to the L3 condition and N = 30 drivers were assigned to the L4 condition.

Table 5.5: Description of study sample, overall and spilt by tested ADF level.

	Total sample	L3 condition	L4 condition
Age	M =38 (SD = 12)	M = 37 (SD = 12)	M = 39 (SD = 12)
Gender	Male: N=32 Female: N=29	Male: N=18 Female: N=13	Male: N=14 Female: N=16
Can do their job while travelling	Yes: N39 No: N=22	Yes: N=20 No: N=11	Yes: N=19 No: N=11
Have a car available for daily use	Yes: N=47 Sometimes: N=8 No: N=6	Yes: N=22 Sometimes: N=6 No: N=3	Yes: N=25 Sometimes: N=2 No: N=3
Driving experience	1-2 years: N=1 2-10 years: N=15 > 10 years:N=45	1-2 years:N=1 2-10 years:N=9 > 10 years:N=21	1-2 years:N=0 2-10 years:N=6 > 10 years:N=24
Frequency of driving	Nearly every day: N=27 3-5 days/week: N=12 1-2 days/week: N=12 less often: N=10	Nearly every day: N=10 3-5 days/week: N=6 1-2 days/week: N=9 less often: N=6	Nearly every day: N=17 3-5 days/week: N=6 1-2 days/week: N=3 less often: N=4
Technology readiness	Among last: N=12 Middle: N=35 Among first: N=14	Among last: N=7 Middle: N=18 Among first: N=6	Among last: N=5 Middle: N=17 Among first: N=8
Have & use ADAS	Parking assist: N=29 Self parking assist: N=2 CC / ACC: N=26 BLIS: N=5 LDW: N=8 LKA: N=4 FCW: N=7	Parking assist: N=9 Self parking assist: N=0 CC / ACC: N=9 BLIS: N=1 LDW: N=3 LKA: N=1 FCW: N=3	Parking assist: N=20 Self parking assist: N=2 CC / ACC: N=17 BLIS: N=4 LDW: N=5 LKA: N=3 FCW: N=4

5.5 Results

In contrast to the general user and acceptance evaluation of the project, the focus of this study was on changes in usage and acceptance resulting from repeated usage of the ADF. For research questions dealing with the change over time, results for the L3 and the L4 condition were combined (unless otherwise stated). For the analysis of objective indicators derived from driver behaviour during experimental drives, only session 1, session 2, session 4 and session 6 were analysed. Session 3 and session 5 comprised the drives for studying driver fatigue with always one of the two drives taking place under sleep deprivation. Therefore, to separate the effects from repeated usage from the effects of variations in drivers' state, drives 3 and 5 were excluded from the analysis of BA.

Additionally, the simulator study addressed two special research questions . Usage and acceptance of an L3 ADF and an L4 ADF were compared and usage and acceptance of the ADFs were also evaluated for drowsy drivers.

If nothing else is stated, graphs in the results section show means and standarddeviations.

5.5.1 Change of willingness to use over time (RQ-U1)

Willingness to use was measured with subjective and objective indicators. Reported willingness to use the ADF increased with repeated usage, significantly so for TJM33a (*I would use this system if it was in my car*, F(3,171)=3.11, p<0.05) and for TJM33p (*I would use the system during my everyday trips*, F(3, 163)=7.59, p<0.001, see Figure 5.2).

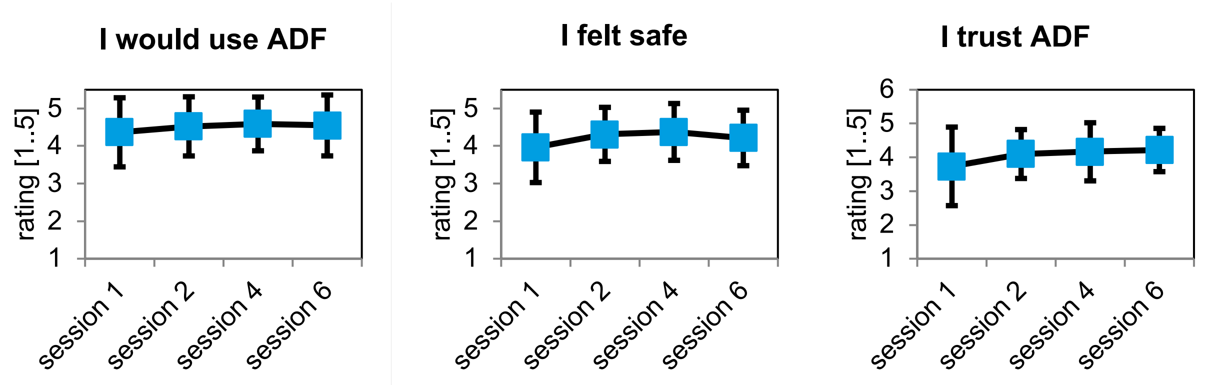

Figure 5.2: Change of willingness to use, experienced safety and trust with repeated usage. 1 = "Strongly disagree" to 5 = "Strongly agree".

There was no change of actual system usage over time, operationalised as proportion of time during which the system was active. This is due to the very high levels of system activation in the 1st session (91% of the time the system was available).

5.5.2 Change of acceptance and trust (RQ-U3)

Acceptance and trust were measured exclusively by questionnaire. There was a significant increase of perceived safety over time (TJM33c, *I felt safe when driving with the system active*, F(3, 165)=9.61, p<0.001), perceived comfort (TJM33q, *Driving with the system active was comfortable*, F(3, 168)=5.27, p<0.01) and of trust (TJM33o, *I trust the system to drive*, F(3, 165)=4.31, p<0.01). Furthermore, results indicate an increase in perceived reliability, reflected in a significant decrease of subjectively unexpected system behaviour (TJM33b, *Sometimes the system behaved unexpectedly*, F(3, 171)=3.60, p<0.05). There is no change over time in the two other items related to perceived reliability (TJM33k, *The system worked as it should work*, F(3, 168)=1.31, p=0.275 and TJM33m, *The system acted appropriately in all situations*, F(3, 168)=1.37, p=0.253).

Furthermore, it was investigated whether the experienced criticality of take-over situations influenced the overall evaluation of the system. For this purpose, trips were divided based on the criticality rating given directly after each take-over situation. Trips were divided into those where maximum experienced criticality was in the range of "harmless" (<=3), "unpleasant" (<=6) and "dangerous" (>6). Ratings of the ADF after the drive changed systematically with experienced criticality of take-over situation during the drive (TJM33c, *I felt safe when driving with the system active*, F(2, 360)=11.96, p<0.001, TJM33q, *Driving with the system active was comfortable*, (F(2, 360)=10.889, p<0.001); TJM33o, *I trust the system to drive*, F(2, 360)=8.94, p<0.001, see Figure 5.3). Consequently, willingness to use was also affected by the criticality of the take-over situation (TJM33a, *I would use this system if it was in my car*, F(2, 360)=32.96, p<0.001; TJM33p, *I would use the system during my everyday trips*, F(2, 360)=13.61, p<0.001).

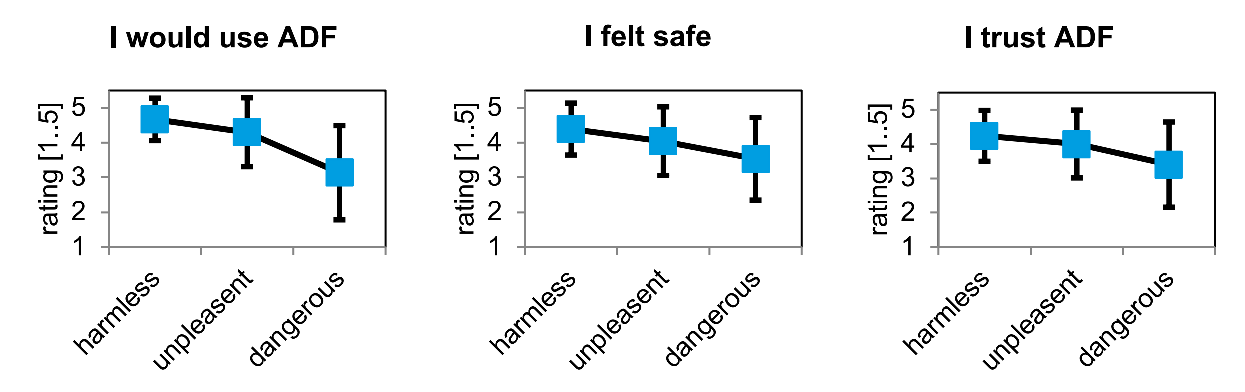

Figure 5.3: Rated willingness to use, experienced safety and trust in relation to experienced criticality of take-over scenarios. 1 = "Strongly disagree" to 5 = "Strongly agree"

5.5.3 With growing experience, understanding of the system increases (RQ-U4)

A total score was calculated for each driver for each driving session based on the questionnaire items assessing the mental model of the systems. Correct answers were scored as 1, incorrect answers as -1 and the option "I don't know" with 0. For the total score of the 20 items a value between -20 and 20 was possible.

There was no significant effect of driving session on the total score (F(3, 178)=0.70, p=0.622). In general, the system understanding was rather high with a mean of 13.63 (SD=3.47). According to the questionnaire data, the understanding of the system did not improve with increasing exposure to the system.

Furthermore, the analysis explored whether drivers learned something about system limits and took control back more frequently before a TOR was issued by the ADF with increasing experience. Separate analyses were conducted for different take-over scenarios (see Figure 5.4).

The exit scenarios occurred at the end of every drive. With L3-ADF which has a rather late TOR (15 seconds before the exit is reached), drivers took control back quite frequently before the TOR. This was presumably because the navigation system informed the driver about the oncoming exit before the TOR of the L3 ADF. For the L4 system, the TOR by the ADF occurred before the information from the navigation system. Here, drivers only rarely took control back before the TOR. For both levels, there was no change in behaviour with repeated usage.

A similar picture was found before the scenario highway crossing, which occurred in the 1st, 2nd and 6th session. There was a considerable difference between the two ADF levels but no change with repeated usage.

The ‘bad lane marking’-scenario was a take-over scenario that occurred only with the L3-ADF and was not signalised by any technical device in the car or by hints in the scenery. In that scenario, control had been taken back before a TOR occurred in only 7 out of 155 instances; again, there was no change with increasing experience with the system.

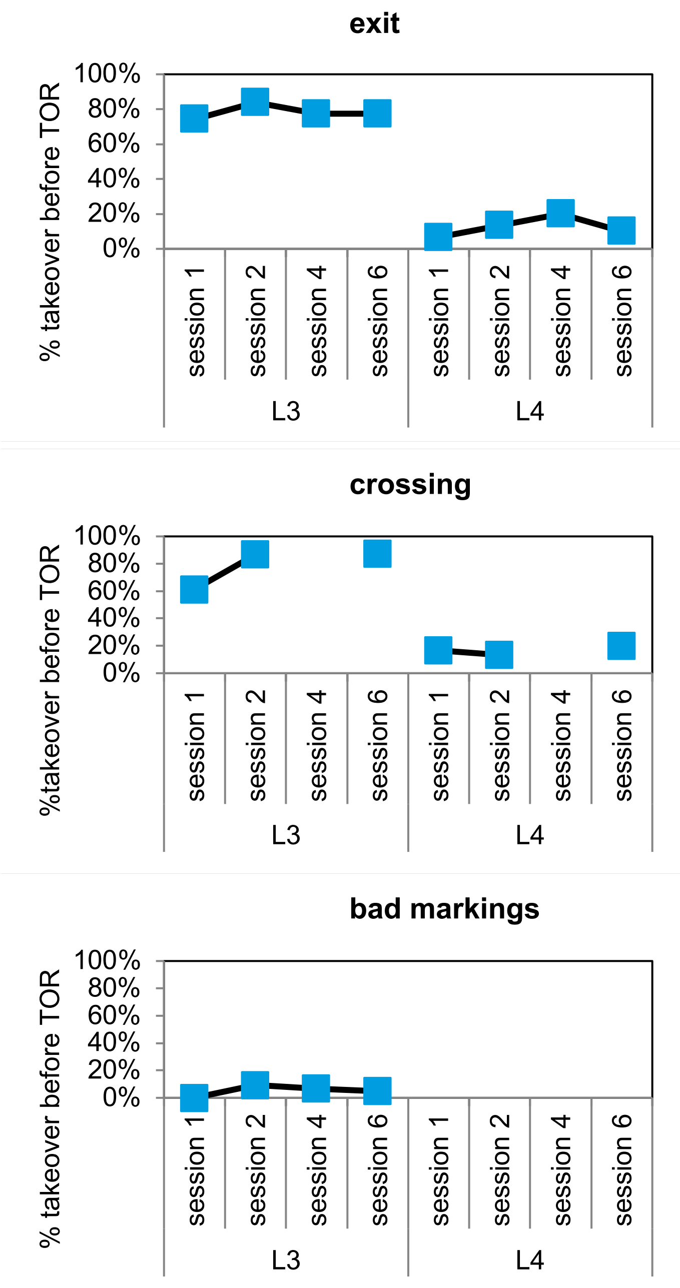

Figure 5.4: Change of frequency of take-over before a TOR with repeated usage split by ADF level and driving scenario. In session 4 there were no TORs due to crossings and with L4-ADF there were no TORs due to bad lane markings.

5.5.4 Change of driver state with repeated usage (RQ-U5)

Over time there was a significant decrease of reported stress (TJM33j, *Driving with the system was stressful*. $F(3, 171)=6.33$, $p<0.001$, see Figure 5.5) and of reported workload (TJM33i, *Driving with this system was demanding*. $F(3, 171)=4.47$, $p<0.01$).

With repeated usage drivers believed less strongly that driving with the ADF would make them tired (TJM33t, *Driving with the function on long journeys would make me tired,* $F(3, 165)=2.84$, $p<0.05$). This finding was in contrast with the absence of change in drivers' state measured via KSS ($F(3, 234)=0.58$, $p=0.63$). Nevertheless, there was an increase of KSS ratings between directly before and after a 20-30 minutes-drive ($F(1, 238)=45.29$, $p<0.001$) of a bit more than half a scale point.

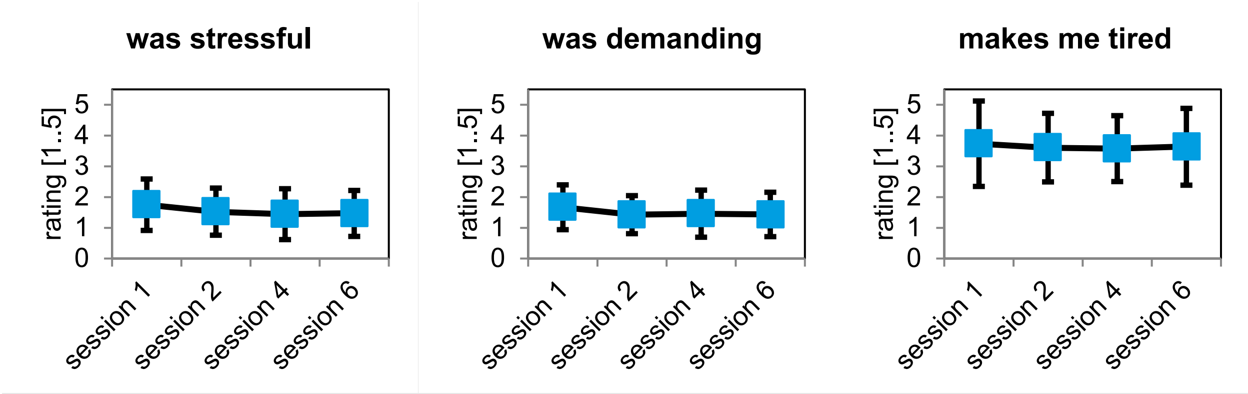

Figure 5.5: Change of experienced drivers state with repeated usage. 1 = "Strongly disagree" to 5 = "Strongly agree"

To get a more objective assessment of fatigue, PERCLOS was derived from the logged eyelid opening level. To be sure that the analysed data was reliable, only sections with manual driving of at least 15 seconds were included in the analysis. Since the number of manually driven sections varied between drives and drivers (mininmum 2 sections, maximum 12 sections), the first and last of the sections were included in the analysis to see whether there was a systematic impact of time and driving session on PERCLOS. The results showed that there was neither a significant impact of time (beginning vs. end of drive) nor an impact of session on PERCLOS.

5.5.5 Change of drivers' attention to other road users over time (RQ-U6)

Attention to the other road users was assessed via questionnaire and the measured proportion of glances directed to the road. There was a significant decrease in how much drivers wanted to monitor the system (TJM33l, *I would want to monitor the system's performance*. $F(3, 168)=5.59$, $p<0.01$), in how much they actually felt that they monitored the environment (TJM33r, *During driving with the system active, I monitored the surrounding environment more than in manual driving*. $F(3, 171)=9.38$, $p<0.001$) and in how much they felt they were aware of hazards (TJM33s, *During driving with the system active, I was more aware of hazards in the surrounding environment than in manual driving*, $F(3, 171)=2.85$, $p<0.05$).

This reported decrease of awareness to the road was supported by a significant decrease in the proportion of glances directed to the road (PRC, $F(3, 180)=8.21$, $p<0.001$, see Figure 5.6) from 25% of time with the system active in the 1st session to 16% in the 6th session. To test whether this reduction of awareness was dependent on the driving situation, the PRC was calculated separately for situations with stable lane bound driving and situations, where the ADF overtook a slower vehicle. During all session, drivers directed less attention to the road during stable lane bound driving compared to during overtaking (F(1, 59)=12.27, p<0.001), and there was a significant decrease over time (F(3, 177)=7.31, p<0.001) but no interaction. This result might indicate that although drivers reduced their attention to the road with increasing experience with the system they still directed more attention to driving in more dynamic situations such as overtaking.

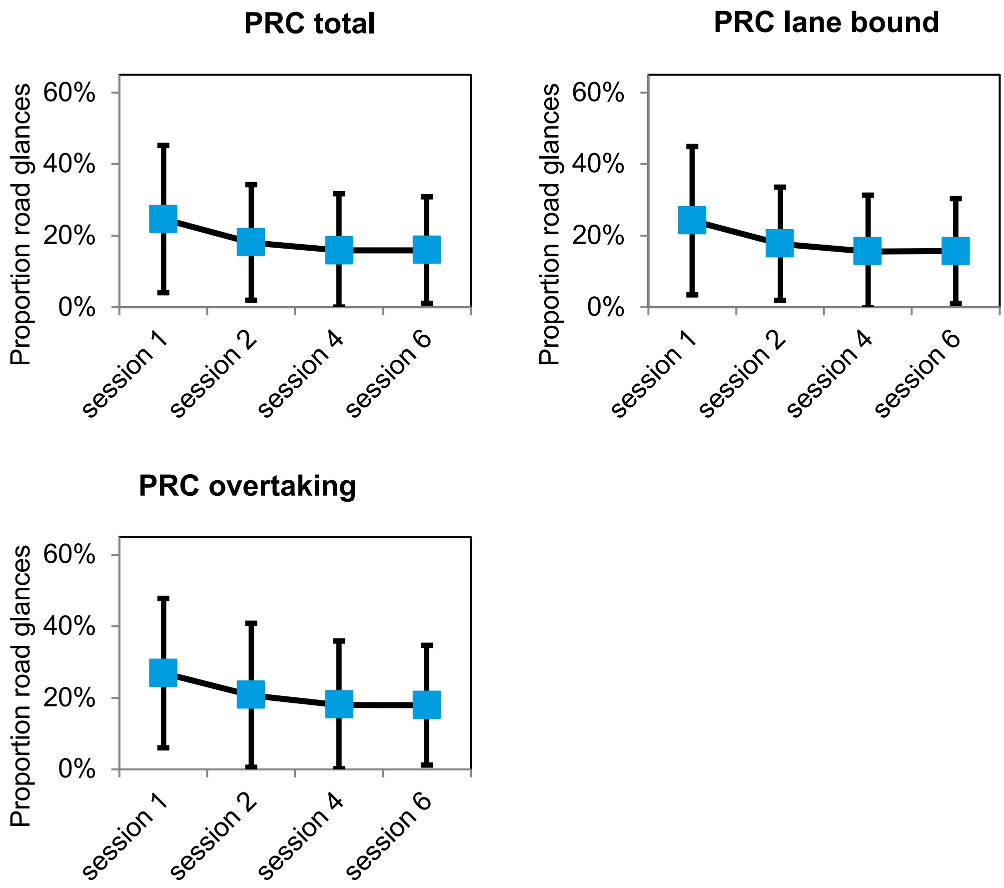

Figure 5.6: Change of percent road glance (PRC) with repeated usage.

5.5.6 Change of secondary tasks engagement with repeated usage (RQ-U9)

There was a significant increase in the extent to which drivers agreed with item TJM33n (*I would use the time the system was active to do other activities*, F(3, 165)=7.24, p<0.001). This was in line with a significant increase of secondary task engagegements with repeated usage (F(3, 180)=10.90, p<0.001) from 59% of time driving with the system active in the 1st session to 73% in the 6th session. This corresponds to a relative increase of 24%.

The increase was especially pronounced for tasks with active involvement of both hands: The proportion of time spent on activities which involved both hands rose significantly from 35% in the 1st to 57% in the 6th session (F(3, 180)=8.91, p<0.001). This equals a relative increase of 63%. This increase in activies involving both hands was found for both systems (L3 & L4), but there was a significant interaction between the effect of repeated usage and system type ($F(3, 165)=4.02$, $p<0.01$). For the L4 ADF, the main increase was observed between the 1st and the 2nd session; for the L3 ADF the main increase occurred later (between 2nd and 4th session).

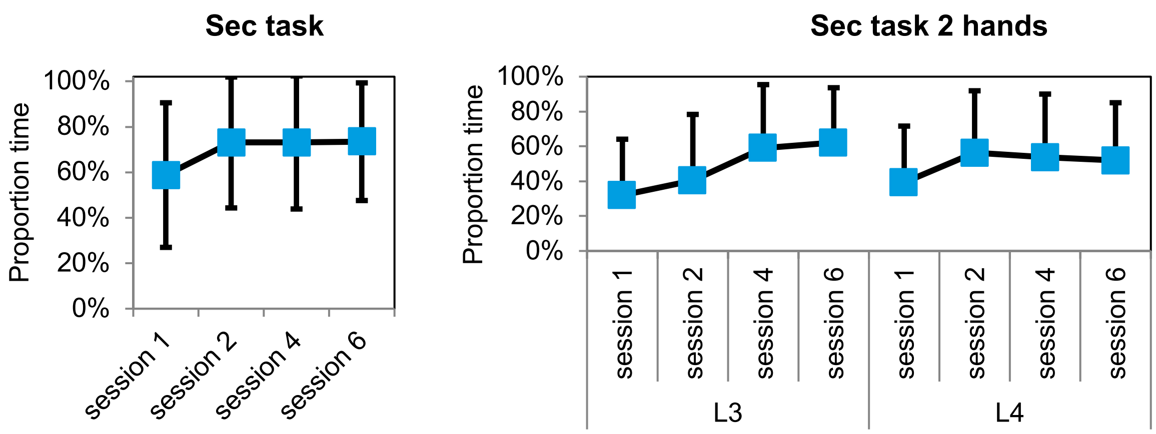

Figure 5.7: Change in proportion of time spent on secondary tasks (left) and secondary tasks which involve both hands (right) while driving with ADF active.

5.5.7 Change of take-over performance with repeated usage (RQ-U10)

Two approaches were used to evaluate take-over performance: First, reaction times for the take-over reaction were calculated. This was the time it took until drivers looked on the road (Eyes On Road time), until drivers put their hands on the steeringwheel (Hands On time) and until drivers deactivated the function (Take-over time). Because there was a large impact the ADF level (L3 vs. L4), the factor ADF-level was included in the analysis of indicators for take-over reactions. For the parameter Eyes on Road time there was only a significant difference between ADF level ($F(1, 372)=26.61$, $p<0.001$). With L4, drivers took longer to look at the road (on average 1904 ms vs. 558 ms) than with L3. For the two other parameters, there were significant main effects for ADF level (Hands On time: $F(1, 509)=127.98$, $p<0.001$, System off time: $F(1, 515)=199.66$, $p<0.001$) and significant interactions (Hands On time: $F(3, 509)=4.74$, $p<0.01$, System off time: $F(3, 515)=3.46$, $p<0.05$). Drivers took significantly longer to put their hands on the steering wheel and to turn off the ADF while driving with L4. Furthermore, with L4 ADF take-over times significantly increased with repeated usage. There was no change over time for the L3 ADF. In summary, it seems that BA occurs for take-over responses but only if the design of the take-over request (e.g. the available time budget) allows such an adaption.

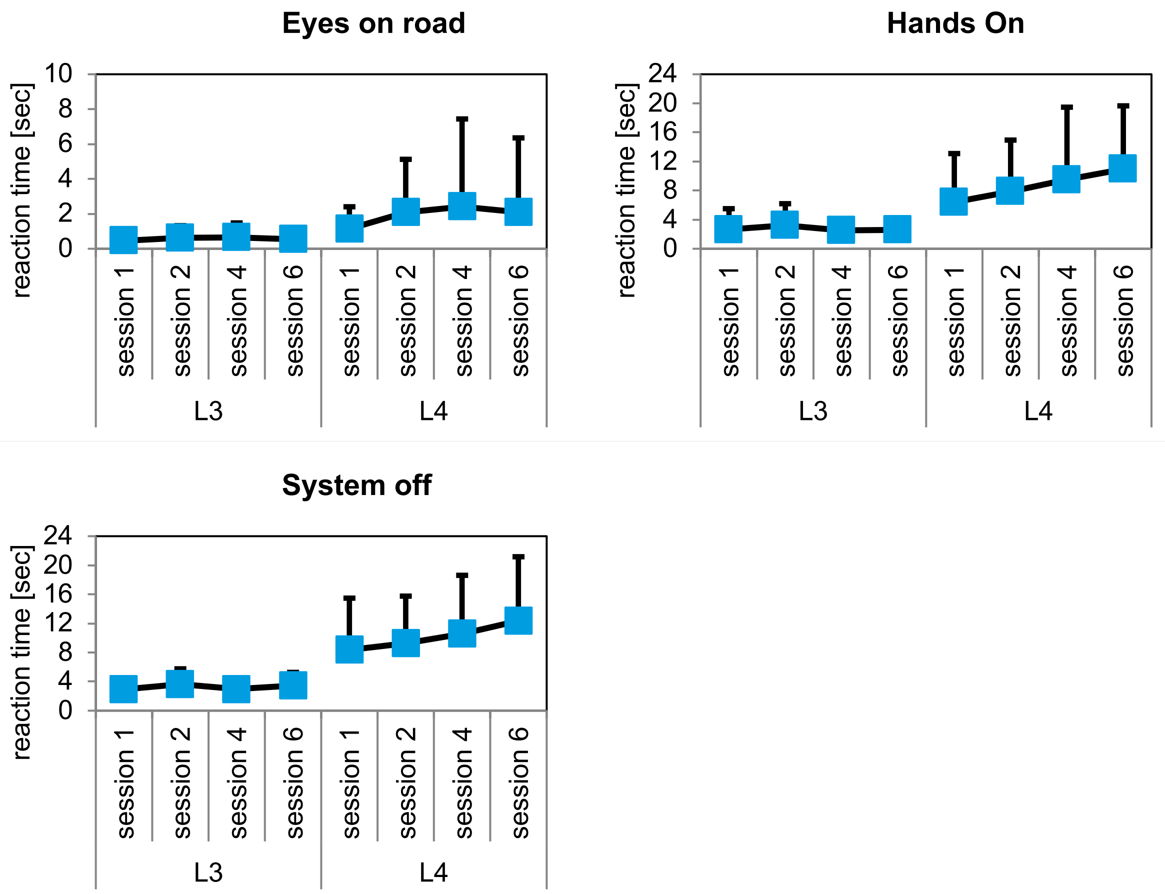

Figure 5.8: Change of reaction times (eyes on road time (upper left), hands on time (upper right) and time until system is turned off (low)) after a TOR with repeated usage, seperatly for L3 and L4 systems.

For the second approach, the overall take-over performance was evaluated with the TOC-rating. For this indicator, there was neither a significant difference between ADF levels nor a significant change over time. Subjectively, drivers experienced take-over situations as less critical with L4 ADF ($F(1, 367)=40.31$, $p<0.001$). There was no change of experienced criticality with repeated usage.

5.5.8 Change of system activation / deactivation with repeated usage (RQ-U11)

To analyse whether the pattern of system activation / deactivation changed systematically with repeated usage, several indicators were analysed:

- Proportion of time with system active in sections with speed limit and free flow conditions (occurring in all four sessions analysed): significant effect of session ($F(3, 180)=4.74$, $p<0.01$) due to lower usage in session 4 (90% of time with system available vs. 96% of time in the other sessions).
- Proportion of time with system active in sections with no speed limit and free flow conditions (occurring in all four sessions analysed): nearly significant effect of session ($F(3, 180)=2.41$,

p=.068) due to lower usage in last session (94% of time with system available vs. 97% of time in the other sessions).

- Proportion of time with system active in sections with traffic jam conditions (occurring in three out of four sessions): nearly significant effect of session ($F(2, 112)=2.80$, $p=.065$) due to higher usage in the 3rd session (99% of time with system available vs. 96% of time in the other sessions).
- Proportion of time with system active while driving in stable conditions on preferred lane (middle lane) and proportion of time with system active in overtaking conditions (driving on left lane): no difference between overtaking yes/no and no change with repeated usage.

In summary, there was no relevant systematic change of the activation / deactivation patterns with increasing experience with the system. It might be that such changes could not be observed in the study because the overall activation of the ADF was high with over 90% of the time where the system was available and because situations especially challenging for an ADF (e.g., higher traffic density with high differences in speed between the lanes) were not included in the drives. Nevertheless, there was a tendency for sections with no speed limit: if drivers drove manually, they drove faster with repeated usage ($F(3, 19)=2.73$, $p=.073$). Average manual speed increased from 123 km/h in session 1 and 2 to 150 km/h in the last session).

5.5.9 Acceptance and usage of L3 vs. L4 (RQ-UE2)

The Acceptance scale (Van Der Laan, 1997) in the L3Pilot questionnaire (TJM.31) was administered after each drive. For the acceptance and perceived usefulness, the Acceptance scale was compared after the 6th drive for the L3 ADF and the L4 ADF. The scale can be divided in two sub-scales, the *Satisfying* scale, and the *Usefulness* scale. Sub-scale scores were calculated according to the instructions in the original paper (Van Der Laan, 1997).

Independent t-tests were calculated to compare the two subscales for the L3 ADF and the L4 ADF. The *Satisfying Scale* received higher rating from participants using the L4 ADF ($M=1.7$, $SD=0.4$) compared to the L3 ADF ($M=1.0$, $SD=0.8$; $t(58)=4.0$, $p=.002$). The same was true for the *Usefulness Scale*: usefulness was rated higher for the L4 ADF ($M=1.3$, $SD=0.4$) than for the L3 ADF ($M=0.7$, $SD=0.8$; $t(58)=3.8$, $p<0.01$).

Customised items of the L3Pilot questionnaire also aimed at user acceptance and evaluation of the ADFs (TJM.33a-ß). The answers ranged from 1 = “Strongly disagree” to 5 = “Strongly agree”. Independent t-tests were calculated for each item comparing the L3 ADF and the L4 ADF. Table 5.6 shows test values of those items that showed a significant difference between the L3 ADF and the L4 ADF. The results support the results found by the standardised acceptance scale. The overall acceptance was high for both systems, however the perceived safety and trust were higher for the L4 system (TJM.33c, TJM.33l, TJM.33o). The L4 system was also perceived as less difficult (TJM.33h) and less stressful to use (TJM.33j). Also the willingness to use (TJM.33a, Tjm.33p) and the willingness to buy (TJM.33d) were higher for the L4 condition.

Table 5.6: Overall evaluation of the ADF split by ADF level.

Item	M(L4)	M(L3)	SD(L4)	SD(L3)	p
I would use this system if it was in my car. (TJM.33a)	4.9	4.2	0.3	1.0	.002**
I felt safe when driving with the system active. (TJM.33c)	4.5	3.9	0.6	0.8	.002**
I would buy the system. (TJM.33d)	4.4	3.9	0.9	1.1	.045*
I would recommend the system to others. (TJM.33g)	4.8	4.1	0.5	0.7	.000**
Driving with the system was difficult. (TJM.33h)	1.1	1.4	0.3	0.5	.014*
Driving with the system was stressful. (TJM.33j)	1.3	1.7	0.4	0.9	.001**
The system worked as it should work. (TJM.33k)	4.7	4.0	0.5	0.9	.037*
I would want to monitor the system's performance. (TJM.33l)	2.9	3.6	0.9	1.0	.003**
I trust the system to drive. (TJM.33o)	4.4	4.0	0.6	0.6	.008*
I would use the system during my everyday trips. (TJM.33p)	4.8	4.4	0.4	0.7	.011*
Using the system on motorways was fun. (TJM.33u)	4.7	4.1	0.6	0.8	.001**

An increased perceived usefulness was also apparent for the willingness to engage in NDRAs (TJM.34). After the 6th driving session, the overall willingness to engage in NDRAs was high for both ADF levels. The most popular NDRAs were "Music, radio, audiobooks", "Interact with a passenger" (both were not available in the simulator study), "Texting" and "Browsing the internet". Differences between ADF-levels could be observed for execution of the NDRAs "Watching a movie" and "Sleeping". These to NDRAs were rated to be carried out more frequently with the L4 ADF compared to the L3 ADF.

Table 5.7: Willingness to engage in different NDRAs split by ADF level.

NDRA	M(L4)	M(L3)	SD(L4)	SD(L3)	p
Music, radio, audiobook	5.87	5.60	0.43	0.67	.074
Interact with a passenger	5.73	5.47	0.58	0.68	.109
Texting	5.33	4.93	0.76	0.98	.082
Browsing the internet	4.97	4.70	0.85	1.24	.334
Navigation	4.80	4.43	1.06	0.90	.154
Calling	4.77	4.23	0.94	1.17	.055
Smart phone apps	4.77	4.53	1.07	1.43	.478
Eating or drinking	4.70	4.93	0.92	0.78	.294
Social Media	4.57	4.07	1.22	1.51	.164
Office/work tasks	3.87	3.73	1.46	1.31	.711

NDRA	M(L4)	M(L3)	SD(L4)	SD(L3)	p
Watching movies	3.70	2.97	1.18	1.52	.041*
Sleeping	3.60	2.20	1.35	1.45	.000**
None	2.37	1.93	1.25	1.20	.175
Personal hygiene/Cosmetics	2.23	2.00	1.30	1.29	.488
Smoking	1.23	1.33	0.90	1.06	.695

*significant on the .05 level

**significant on the .005 level

5.5.10 Impact of AD level on driver state (RQ-U5)

Drivers agreed with the statement that driving with the function would make them tired already after the 1st drive. There was no impact of AD-level on that result. This result is supported by KSS-ratings provided before and after every trip. The change in fatigue experienced as measured by the KSS is neither dependent on ADF-level nor on increasing experience with the ADF.

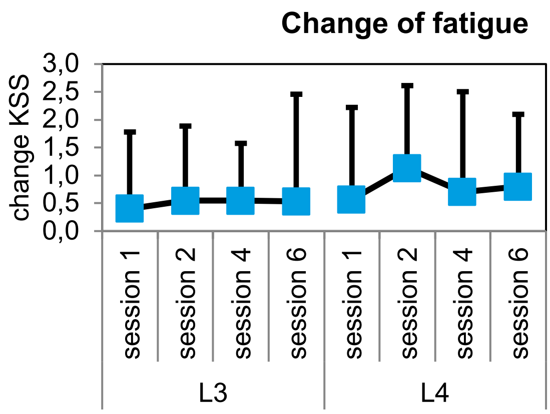

Figure 5.9: Impact of ADF-level and repeated usage on the change of KSS-ratings over one drive.

5.5.11 Usage and acceptance of AD by fatigued drivers (RQ-UE3)

To evaluate whether driver fatigue has an effect on acceptance of the ADF, subjective ratings collected after the last drive were compared with ratings following the two monotonous drives, one with sleep deprivation (fatigued drivers) and one without sleep deprivation. That drivers were more fatigued after sleep deprivation was reflected in KSS ratings gathered directly before the start of the drive (F(1, 57)=249.4, p<0.001, m(not deprived)=3.6, m(deprived)=7.2) as well as in PERCLOS level measured during a ten minute manual driving section directly at the beginning of the not-deprived drive and once during the drive with sleep deprivation (F(1, 55)=7.64, p<0.01, m(not deprived)=4.1%, m(deprived)=8.3%). Both, subjective and objective measures of fatigue indicated increased fatigue in the session with sleep deprivation.

For items reflecting overall evaluation of the ADF (e.g. TJM.33c, TJM.33l, TJM.33o, TJM.33a, Tjm.33p) there was a significant difference between ADF levels (see RQ-UE2) but no impact of drivers' state. This was probaby due to the highly positive evaluation of the ADF overall which was also reflected in a very high usage (over 95% of time the ADF is available).

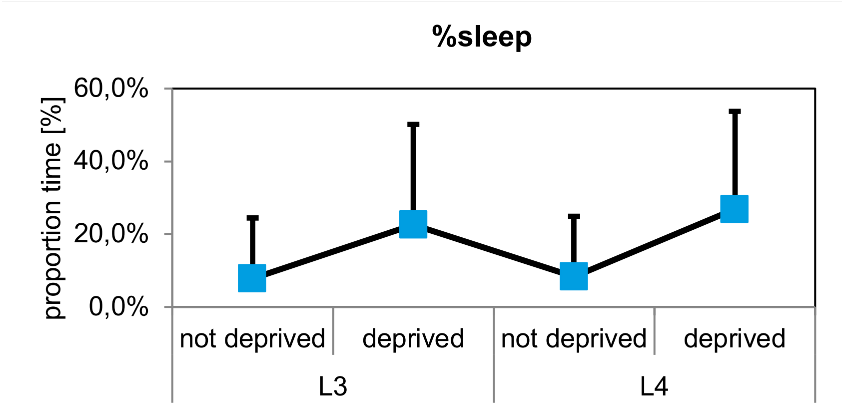

Figure 5.10: Proportion of time spent sleeping for drivers with and without sleep deprivation and for the different ADF levels.

Sleep while driving with the ADF active was measured based on EEG-recordings logged during the two monotonous driving session. Proportion of sleep includes driving sections that were coded as beginning sleep (N1), stable sleep (N2) or deep (N3) based on the AASM clinical standard (AASM, 2017). Independent of system level, fatigued drivers (that is after sleep deprivation) used the time driving with the ADF to a larger proportion to sleep than non fatigued drivers ($F(1, 117)=16.57$, $p<0.001$, see Figure 5.10).

5.6 Summary of results

Taking all results on behavioural adaption together, there was an increasingly positive evaluation of the ADF with increasing experience with the system. The subjectively reported increase in trust was reflected in system usage: drivers spent more time on side tasks and directed less attention to the road. In addition, experienced stress decreased and driving with the ADF became more comfortable and less demanding. Contrary to expectation there was neither a significant change of pattern of system activation and deactivation (drivers kept the system activated more than 90% of the time in all trips) nor a change of system understanding. Furthermore, with increasing experience with the system drivers did not get more fatigued driving with the ADF engaged.

In situations where the ADF requested the driver to intervene, control was taken back safely. There was neither an impact of increased experience with the ADF nor of system level on the take-over performance. Contrary to that, take-over times increased with growing experience with the ADF; however, only for the L4 implementations. It seems that drivers learn with increasing experience with the ADF that they can use the available time budget to take control back safely, and that there is no need for a quick reaction. This was only the case for the L4 implementation because only here was the time budget provided by the ADF sufficient to allow behavioural adaption. For the L3 ADF, the available time budget of 15 seconds was probably not enough to allow systematic adaptation of take-over reactions.

Although investigated as a between group factor, the L4 ADF was evaluated more positively on many dimension like satisfaction and usefulness, willingness to use, trust, experienced safety etc. After getting to know the L4 ADF, drivers stated that they would use driving time more frequent to watch movies or to sleep than drivers testing the L3 ADF. However, this was not in line with the results from the trips with fatigued drivers. Independent of ADF level, fatigued drivers used on average about 25% of driving time to sleep.

6 On-road study on long-term behavioural adaptation

This chapter describes an on-road study that explored the change of acceptance and usage of an L4-ADF with growing experience. The study was conducucted by Renault, and the analysis was done by WIVW and Leeds University.

6.1 Aim and research questions

The following results are based on data collected in an on-road pilot study. Drivers participating in the study had the chance to test an L3-motorway chauffeur three times. System evaluation was assessed using the L3Pilot questionnaire after the first and third drive. This data was used to assess changes of system acceptance and evaluation with repeated usage. Here, the focus was on common L3Pilot RQs, but with a specific focus on the change of the investigated topics with repeated usage of an ADF. Specifically, the following RQs were addressed:

- RQ-U1: Are drivers willing to use an ADF?
- RQ-U3: What is the user acceptance of the ADF?
- RQ-U5: What is the impact of ADF on driver state?
- RQ-U6: What is the impact of ADF use on driver awareness?

The L3Pilot questionnaire (Metz et al. 2020) was filled in by participants twice, after 1st contact with an ADF and after 3rd usage.

6.2 Method

6.2.1 Automated vehicle and route

The study was conducted from January 2020 to March 2021 and took place on a 95km long motorway section close to Paris, France (see Figure 6.1).

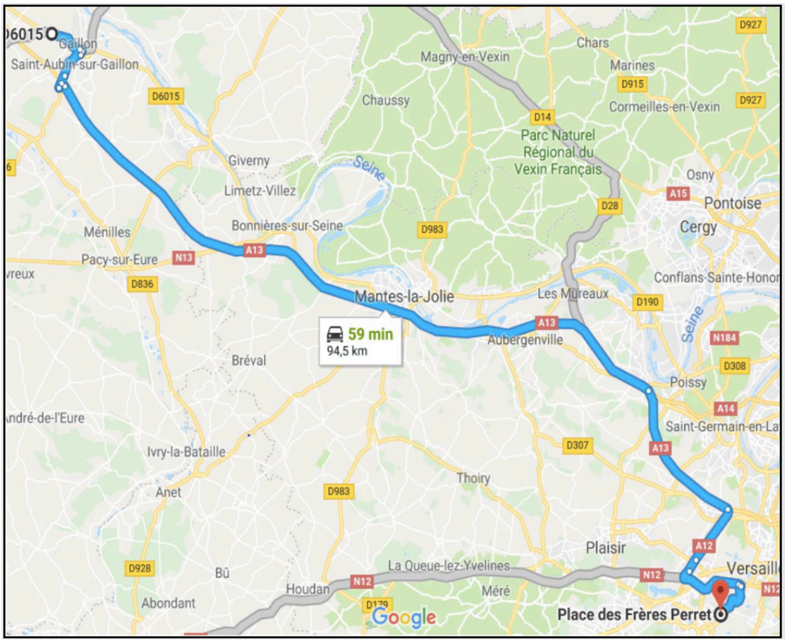

Figure 6.1: Motorway section, including start- and end-point.

The experimental demonstrator vehicle was a Renault Espace provided by Renault France. The vehicle's Automated Driving Function (ADF) became available on the motorway, and included driving on its lane, performing overtaking manoeuvres, and changing lanes. For safety reasons, the AV's system did not perform the lane-changes automatically, and these were initiated by the safety driver but executed by the vehicle. Another manually driven vehicle followed the AV on the motorway to ensure safe lane changes for the AV.

6.2.2 Experimental Procedure and Design

This study consisted of three automated drives on the motorway, lasting approximately 1 to 1.5 hours, depending on the traffic flow. The second drive was conducted around two to three weeks after the first one, and the third drive around two to three months after the first one.

Before the first experimental drive, all participants were debriefed about the experiment, and given an opportunity to ask questions. They also received an informed consent form and a pre-experimental questionnaire (see Section 6.2.3). Prior to each experimental drive the participants were given the opportunity to do a practice drive (3-4 km on a rural road) to become familiar with the vehicle and the ADFs. Initially, they were given options on whether they needed the practice drive before the second and third drive, and some participants did not need another practice drive. However, it became mandatory for them to do a practice drive due to the time gap between drive two and three caused by the Covid-19 pandemic-related interruption of the experiment (two breaks: from March to June and from November to December 2020).

At the beginning of each experimental drive, participants drove to the motorway manually. Upon arrival on the motorway, the automated driving mode became available if the following three criteria were all fulfilled: The AV

(1) was located in the centre of the lane,

(2) had a certain security distance to the leading vehicle, and

(3) was driving less than 110 km/h.

Only then, the dashboard turned blue, and the message 'the vehicle is ready for Automated Mode' was presented. If these criteria were not fulfilled, the experimenter instructed the participant to adjust the missing parameters. To hand over the driving task, the participant was asked to release the acceleration pedal at first and then to push the button 'R' on the steering wheel. Once activated, the dashboard turned golden, and a sound was provided, which intended to inform the participant that they had activated the automated driving mode. During automated driving, participants were allowed to do what they wanted after handing over the driving task. They could take back manual driving whenever they wanted to or when the AV's system asked them to do so. The participants were prompted to take-over one minute before the motorway exit, or 10 seconds prior to an unexpected event – in these situations the message 'You have XX s to take over control' was displayed on the dashboard and a sound cue was presented. To take over the driving task, the participants had to press the button 'O' on the steering wheel, or press the acceleration pedal, or turn the steering wheel.

For the three experimental drives, the participants received different instructions regarding the duration of driving in automated mode and their secondary tasks during automated driving:

- In the first experimental drive, the participants had the chance to discover the ADFs and to drive automated on the motorway. Participants were instructed to hand over control as soon the automated driving mode was available, but they were always free to take over if they wanted to. During automated driving, they were free to engage in other, non-driving related, activities.
- During the second experimental drive, the participants were asked to drive one half of the motorway section manually and to activate the automated mode during the other half of the drive. There was no instruction regarding a secondary task for the period of automated driving.
- In the third experimental drive, they were instructed to drive automated as soon as the automated driving mode was available, and they were offered to engage in secondary tasks such as reading a book or playing on a smartphone. Participants then completed the final questionnaire, after which they were interviewed and asked how the Traffic Jam Motorway Chauffeur had influenced their behaviour and how they had learned to use this ADF.

6.2.3 Questionnaires

Before the first experimental drive, all participants were informed about the experiment, and given an opportunity to ask questions. Participants were asked to complete a pre-experiment questionnaire, which consisted of 109 items, including questions on driver age, gender, employment, education, Attitude towards technology (6 item scale), and Sensation Seeking (8 item scale).

In addition to the pre-experiment questionnaire, participants completed another questionnaire on two occasions - immediately after the first and third drive. This questionnaire consisted of 81 questions and included items measuring acceptance of automation (van der Laan et al., 1997), perceived comfort of the motorway system's behaviour, willingness to use the automation, perceptions of the automated system, and attention and awareness during automation.

6.2.4 Sample description

A total of N = 80 drivers took part in the study. Table 6.1 gives an overview of the sample.

Table 6.1: Description of study sample.

	Sample
Age	M = 44 (SD = 12)
Gender	Male: N=55 Female: N=25
Can do their job while travelling	Yes: N=53 No: N=27
Have a car available for daily use	Yes: N=69 Sometimes: N=5 No: N=6
Driving experience	2-10 years: N=15 > 10 years:N=65
Frequency of driving	Nearly every day: N=65 3-5 days/week: N=8 1-2 days/week: N=5 less often: N=2
Technology readiness	Among last: N=5 Middle: N=44 Among first: N=31
Have & use ADAS	Parking assist: N=41 Self parking assist: N=2 CC / ACC: N=55 BLIS: N=11 LDW: N=12 LKA: N=5 FCW: N=13

6.3 Results

To evaluate whether system evaluation by drivers changes with repeated usage of the system, anwers to the questionnaire items after the first and third drive were compared. If nothing else is stated graphs in the results section show means and standard deviations.

6.3.1 Change of willingness to use over time (RQ-U1)

Willingness to use was measured with subjective indicators. There was no change of reported willingness to use the ADF with repeated usage for TJM33a (*I would use this system if it was in my car*, F(1,77)=0.07) and for TJM33p (*I would use the system during my everyday trips*, F(1, 77)=0.44). For both items, users' assessments were highly positive after 1st usage already (4.0/4.6 on a 5-point scale), leaving little room for further improvement.

6.3.2 Change of acceptance and trust (RQ-U3)

Acceptance and trust were measured by questionnaire. There was no change over time of perceived safety (TJM33c, *I felt safe when driving with the system active,* F(1, 77)=0.33), perceived comfort (TJM33q, *Driving with the system active was comfortable*. F(1,77)=0.02) and of trust (TJM33o, *I trust the system to drive*, F(1, 77)=0.06). Furthermore, results indicate no change of perceived reliability (TJM33b, *Sometimes the system behaved unexpectedly*, F(1, 77)=1.07; TJM33k, *The system worked as it should work*. F(1, 77)=0.52, and TJM33m *The system acted appropriately in all situations*. F(1, 77)=0.89). For perceived safety, comfort, and trust, users' evaluation was already highly positive after 1st usage (between 4.1 and 4.4 on a 5-point scale), leaving little room for further improvement (see Figure 6.2).

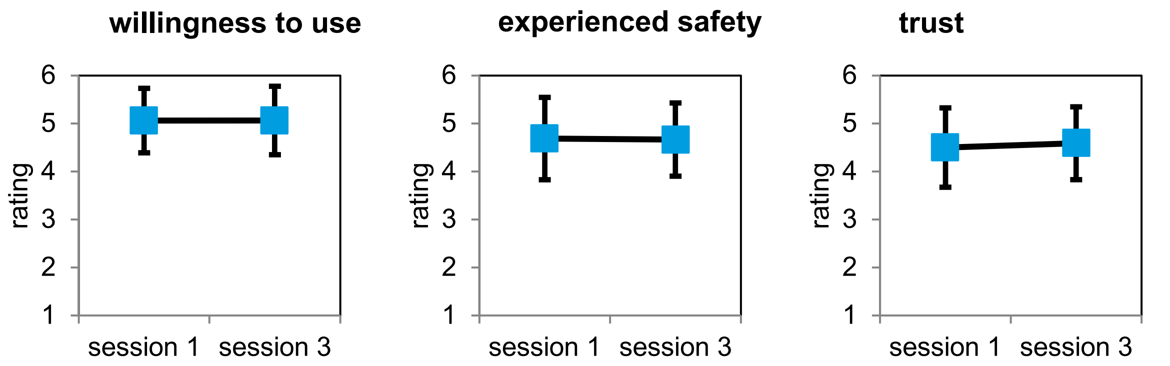

Figure 6.2: Change of willingness to use, experienced safety and trust with repeated usage. 1 = "Strongly disagree" to 5 = "Strongly agree".

6.3.3 Change of driver state with repeated usage (RQ-U5)

Over time, there was no change of reported stress (TJM33j, *Driving with the system was stressful*. F(1, 77)=1.65) and of reported workload (TJM33i, *Driving with this system was demanding*. F(1, 77)=0.61), see Figure 6.3. There was no change with repeated usage in how strongly drivers believed that driving with the ADF would make them tired (TJM33t, *Driving with the function on long journeys would make me tired*, F(1,77)=0.21).

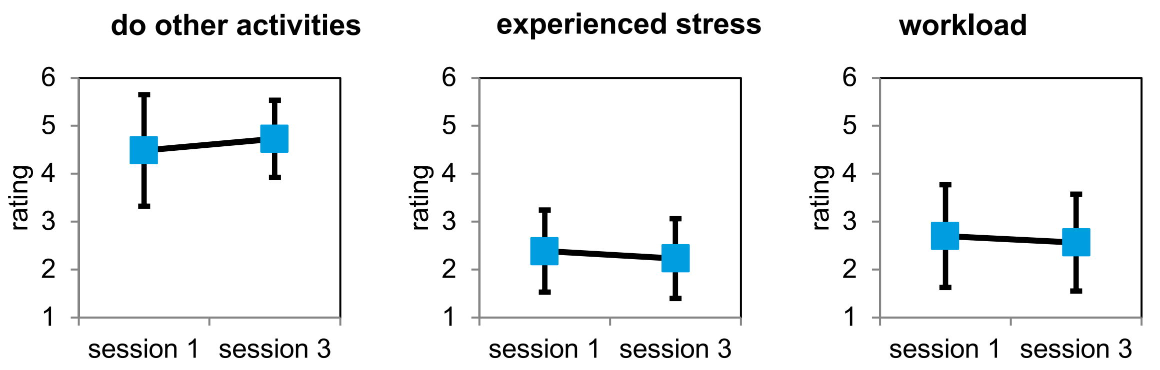

Figure 6.3:Change of willingness to engage in other activities, experienced stress and workload, with repeated usage. 1 = "Strongly disagree" to 5 = "Strongly agree".

6.3.4 Change of drivers' attention to other road users over time (RQ-U6)

Attention to other road users was assessed by questionnaire . There was no change in how much drivers wanted to monitor the system (TJM33l, *I would want to monitor the system's performance.* F(1, 77)=0.00) and in how much they actually felt that they monitored the environment (TJM33r, *During driving with the system active, I monitored the surrounding environment more than in manual driving*. F(1, 77)=0.66; TJM33s, *During driving with the system active, I was more aware of hazards in the surrounding environment than in manual driving,* F(1,77)=2.11), see Figure 6.4.

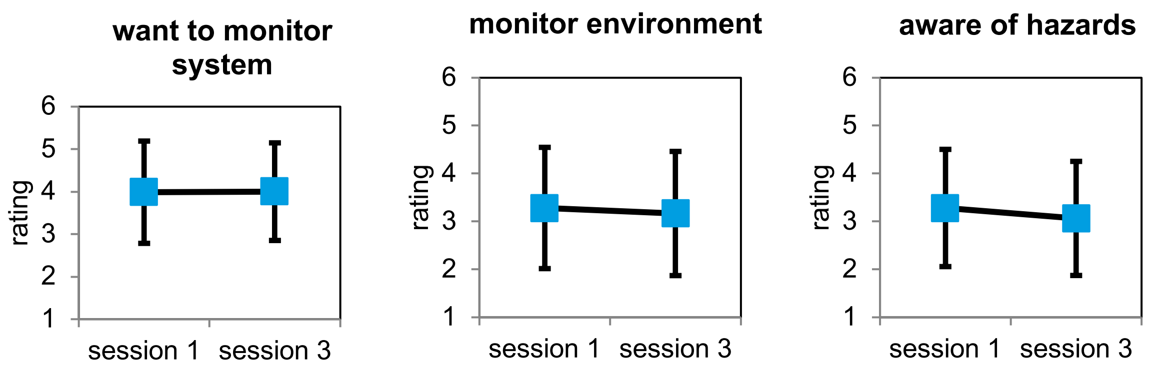

Figure 6.4: Change of items related to drivers' awareness with repeated usage. 1 = "Strongly disagree" to 5 = "Strongly agree".

6.4 Summary of results

Overall, no changes of drivers' attitudes towards the tested ADF could be found in the on-road study. The main reason for this is presumably the highly positive evaluation of the ADF already after the 1st drive. Due to this ceiling effect, no further increase of acceptance was possible. The results clearly show that the tested ADF was accepted by drivers right from the very beginning, and this positive experience was not impacted by actual usage. Drivers experienced driving with the ADF positively even after repeated usage.

7 Wizard of Oz studies on take-over performance and conflict response

Chapter 7 presents the findings from three different test track studies and pilots on public roads. The studies were planned by Chalmers and VCC, the data was collected by VCC staff. All studies were performed with a Wizard of Oz vehicle. Details about these studies are shown in Table 7.1. The results focus on the transition of control from automation to manual driving (take-over).

Table 7.1: The four studies presented in Chapter 7.

	ADEST study	TJP study	L3Pilot Test track study	L3Pilot WoZ pilot
Test environment	Test track	Test track	Test track	Public road
Driver support system/Automation level	L2	L3	L3	L3
Conflict scenario	Lead-vehicle cut-out + stationary object	Road-works zone	Lead-vehicle cut-out + stationary object	None
Conditions	Hands on wheel requirement (yes/no)	Automation duration (4 min/14 min)	Take-over request timings (9 s/18 s time-to-collision)	Repeated exposure to take-over requests
Related chapter	Chapter 7.3	Chapter 7.4	Chapter 7.5	Chapter 7.6

7.1 Aim and research question

The aim of the analyses presented in this chapter was to investigate the drivers' response process when they were required to resume manual control from L3 automation (L3Pilot RQ-U10). The test-track studies investigated take-over performance in conflict scenarios in a controlled environment. In addition, take-over performance in normal (non-conflict) traffic scenarios was investigated in a public road study. Furthermore, the influence of trust on the conflict response was investigated in the ADEST study.

7.2 General Method

7.2.1 The test environments

Investigating take-over responses in conflict scenarios requires a controlled setup both for precise situation replication and to ensure the safety of the test participants. A test track is suitable for this purpose. Take-overs under non-eventful driving can however be studied in real traffic; in this case during the WoZ pilot on public roads.

7.2.1.1 Test track

All test track studies were performed on ASTA Zero rural road test track located in the Gothenburg area (ASTA Zero, 2021). The track is designed to resemble a rural road with a posted speed of 70 kph. There are two travel lanes and the only road users or objects present are the ones specified as part of the experimental design. One lap on the test track is 5.7 km long.

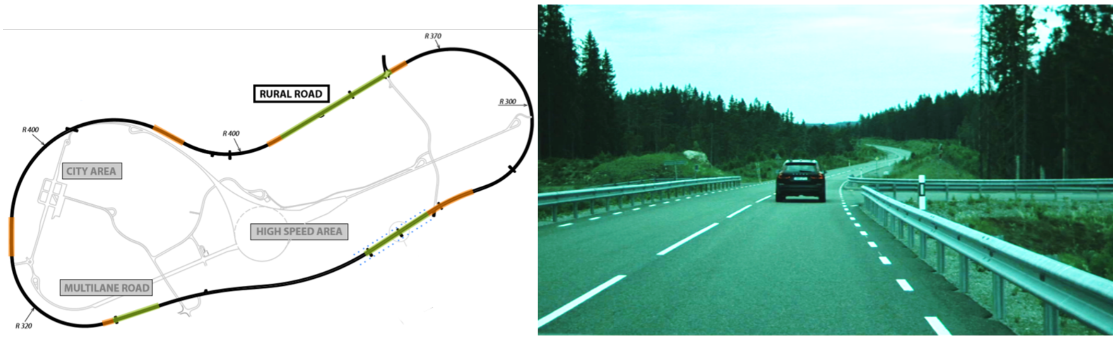

Figure 7.1: ASTA zero rural road test track map (left), and a snapshot from the forward-facing camera on a straight road segment showing a lead vehicle (right).

7.2.1.2 Public road

The Gothenburg ring road was selected for ADF evaluation on public road in the L3Pilot WoZ pilot. The selected route consists of the outer part of the ring-road, illustrated by a dashed line in Figure 7.2. One lap is approximately 30 km long, and the posted speed is 70 or 80 kph. The road is mainly dual carriageway with 2-3 lanes in each direction, separated by median barriers. There are several tunnels and a bridge on the route. Traffic is mostly moderate but gets dense during rush hours.

Figure 7.2: A map of the Gothenburg ring road (left) where the dashed line represents the road segments selected for ADF evaluation in real traffic. A screenshot from the forward view camera (right) shows a straight road segment while travelling south on the eastern part of the ring-road.

7.2.2 The Wizard of Oz test vehicle

The ADF function implemented in the test vehicles used in the main pilot has not yet reached the maturity required for in-production vehicles. These vehicles therefore need to be driven by trained professional drivers and are not suited to address user-related research questions. To overcome this limitation, a Wizard of Oz (WoZ) vehicle was used to enable non-professional drivers to experience and interact with a simulated ADF.

The WoZ vehicle was used both on public roads and the test track described in section 7.2.1. A wizard driver is seated in mid position in the rear seat with access to a steering wheel and a set of pedals. The wizard driver controls the vehicle when in automation mode by driving manually or by using a driver assistance system. The test participants cannot see the wizard's steering wheel and pedals, even though the head and shoulders of the wizard driver are visible from the front seat. The role of the wizard driver is explained to the participant to be that of a safety-driver who will supervise the automation but only intervene if needed.

The participant is seated in the driver's seat and controls the vehicle as in a regular car when the automation is turned off. The wizard driver (or the test leader) can initiate an automation offer when in manual mode, and the participant is invited by an audio tone and a message in the instrument cluster behind the steering wheel to activate the automation. The control is transferred to the wizard driver when the participant presses two buttons on the steering wheel for at least 0.6 seconds. While in automation mode, the wizard driver can issue a take-over request that will provide an audio tone and a message in the DIM requesting the driver to retake control and drive manually. The driver then needs to press the two buttons on the steering wheel to regain control and drive manually. Different degrees of automation can be simulated by changing the instructions to the participants, the presence or absence of take-over requests, or the presence or absence of ADF (wizard) conflict avoidance manoeuvres.

7.3 Driver conflict response in L2 automation (ADEST study)

The ADEST study was performed on the ASTA zero rural-road test track using the WoZ vehicle. The complete study was first reported in Victor et al. (2018). Additional analysis on the response process performed within the L3Pilot project are summarised in this section (see Pipkorn, Victor, Tivesten & Dozza, 2021a for a detailed description). The first aim of this study was to investigate how drivers' conflict response while in supervised automation differed between drivers that crashed with, and drivers who avoided, an on-road object in a conflict scenario. The second aim was to understand the influence of three specific factors on the drivers' response process: a hands-on-wheel requirement (with vs. without), the conflict object type (garbage bag vs. stationary vehicle) and the level of trust in automation to handle the conflict (high vs. low).

7.3.1 Methods

Seventy-six participants supervised a near-perfect L2 automation system (simulated using the WoZ vehicle). Participants followed a lead-vehicle for 30 minutes on the test track before encountering a conflict scenario. In the scenario, the lead-vehicle performed a cut-out to avoid a stationary in-lane object (a stuffed garbage bag or a balloon vehicle). The revealed object was not identified by the automation system, which could therefore not give any feedback or warning to the drivers about the need to intervene. Participants therefore had to act themselves to avoid crashing with the object. After the drive, all participants were asked to what extent they trusted the L2 automation system to handle the conflict scenario they had experienced on a Likert scale from 1 to 7. Scores from 1-3 were considered “low trust”, 4 was considered “mid trust” and scores from 5-7 were considered “high trust”. The drivers’ conflict response was assessed through the response process quantified by the times, relative to the time of passing or hitting the conflict object, for the following driver actions: surprise reaction, hands on wheel, start of steering, and start of braking.

7.3.2 Results

One-third of drivers crashed during the conflict scenario, independently of conflict object type or hands-on-wheel requirement. Crashers generally responded later in all actions of the response process compared to non-crashers. A hands-on-wheel requirement did not influence driver’s conflict response: the drivers with and without hands on the wheel started steering to avoid the conflict object at similar times. In Figure 7.3, the two curves for driver steering with and without a hands-on-wheel requirement almost perfectly overlap. High-trust drivers generally responded later than the low-trust drivers, and only high-trust drivers crashed. The larger object (the balloon vehicle) triggered an earlier surprise reaction compared to the garbage bag, while hands-on-wheel and steering response were similar for the two conflict object types.

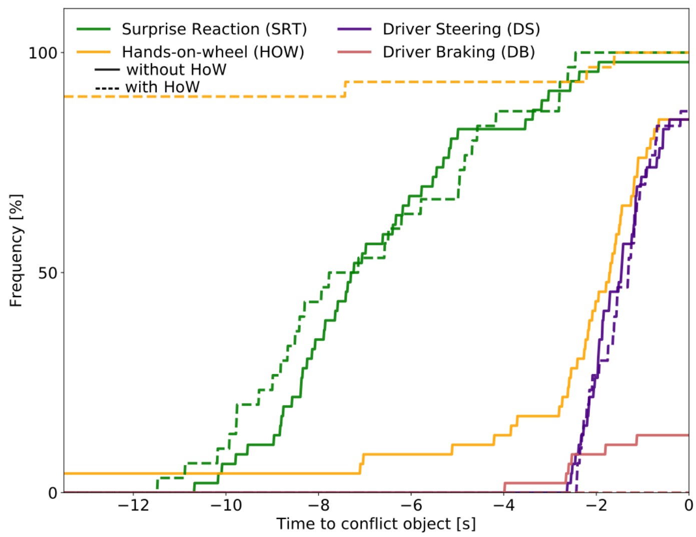

Figure 7.3: The response process for drivers with (dashed line) and without (solid line) a hands-on-wheel requirement. Reprinted from “Driver conflict response during supervised automation: do hands on wheel matter?”, by L. Pipkorn, T. Victor, M. Dozza, & E. Tivesten, 2021a, Transportation Research Part F: Traffic Psychology and Behaviour. Copyright 2020 by Rightslink Inc.

7.3.3 Discussion and conclusions

The results of this study showed that some drivers responded late (put hands on wheel late, start steering late) when supervising a near-perfect L2 automation system in a test track environment, and thus ended up colliding when a conflict object that the system did not recognise was revealed. Also, in this study, a hands-on-wheel requirement did not prevent these drivers from responding late and crashing. However, the extent to which these results can be generalised beyond the test track environment remains unknown, as does the extent to which these results might generalise to other types of conflict scenarios (e.g. sideswipes, lane exits). Further studies are needed before any conclusions that might apply to regular traffic environments can be drawn. Another open question is to what extent drivers would crash to the same extent in a more capable automated system (e.g. L3) that can issue take-over requests prior to the conflict scenario.

7.4 Effect of automation duration on take-over response (TJP study)

The TJP study was performed on the ASTA Zero rural-road test track using the WoZ vehicle. A summary of the study and the findings is presented in this section (see Pipkorn, Victor, Dozza & Tivesten, 2021b for a detailed description). This study was motivated by previous findings of delayed conflict-response observed in driving simulators (e.g. see Gold et al., 2013) and the previously observed crash rates during supervised L2 automation in the ADEST study (see 7.3). This study was performed to understand if drivers would respond late or crash to the same extent as in previous studies in a less critical scenario encountered after a period of L3 automation on test track. The aim of this study was to examine the effect of automation exposure and its duration on the *driver's take-over response* and *driving performance* in an artificial road-works zone. In addition, by comparing the present study's results with previous driving simulator studies, this study also aimed to better understand the influence of factors such as test environment and experimental protocols on the automation aftereffects (i.e. poor manual driving performance after being in automated mode).

7.4.1 Methods

Seventeen participants took part in the study which lasted for 30 minutes. The WoZ vehicle was used to simulate a L3 automation Traffic Jam Pilot (TJP) system with a varying speed profile (up to maximum 70 km/h). During the test, the WoZ vehicle followed a lead vehicle at all times. The participants were instructed to play a game on a tablet mounted on top of the centre stack while TJP was activated. They encountered a road-works zone three times: while driving manually (*manual* condition), and after a short, 4.5 minutes (*AD short* condition) and a long, 14 minutes (*AD long* condition) duration of automation (see Figure 7.4). The order of the short and long duration was counterbalanced: 9 participants experienced the short duration before the long, and the remaining 8 participants experienced the long duration before the short (see Figure 7.4 right). A take-over request (TOR) was issued 5-6 s before the lead vehicle performed a lane change and the road-works zone became visible to the drivers. In response to the TOR, the participants had to press two buttons on the steering wheel to deactivate the TJP. Then, they had to perform manual driving to pass through the road-works zone.

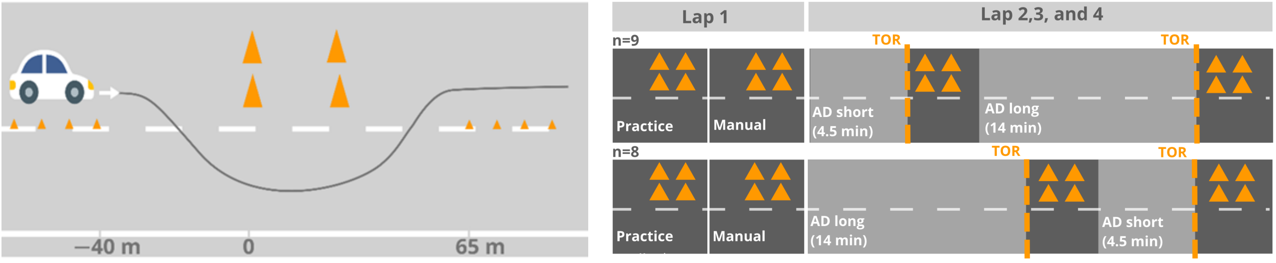

Figure 7.4: The simulated road-works zone built of cones (left), and the study design (right). Reprinted from "Automation Aftereffects: the influence of automation duration, take-over request and timings", by L. Pipkorn, T. Victor, M. Dozza, & E. Tivesten, 2021b, IEEE Intelligent Transportation Systems. Copyright 2020 by Creative Commons CC-BY-NC-ND 4.0.

The driver behaviour in response to the TOR and in the road-works zone was classified as a *driver take-over response* and *driving performance*. The driver take-over response was quantified by the take-over time: the time from the TOR until automation was deactived. The driving performance was quantified by: (a) the time-to-collision (TTC) when drivers started their steering manoeuvre in response to the road-works zone, and (b) vehicle signals for the vehicle speed, longitudinal- and lateral accelerations and steering wheel angle. Take-over time and TTC at driver steering start were modelled using Bayesian varying-intercept models. Both variables were log-transformed and then modelled using a normal distribution. The models were used to quantify effect sizes (i.e. difference in mean times) for AD long vs. manual, AD short vs. manual and AD long vs. AD short. The output of these Bayesian analyses was a posterior distribution for each effect size. The distributions for the backtransformed parameters (i.e. on original scale) were summarized with a mean and a 95% Highest-posterior-density (HPD) interval. The posterior distribution represents the degree of belief in parameter values (e.g. median), and the 95% HPD spans the 95% most probable values (Kruschke & Liddell, 2018).

7.4.2 Results

7.4.2.1 The driver take-over response

On average, the long automation duration resulted in increased take-over times compared to the short duration (see Figure 7.5). The average difference in mean take-over times for a long and a short automation duration was 0.52 s (95% HPD [0.065, 0.95]). The increase in take-over times for the long automation duration compared to the short was influenced by the increased number of failed take-over attempts after the long automation duration compared to the short. In fact, four drivers in AD long required an additional button press to successfully deactivate automation, since the first press was too short (< 0.6 s).

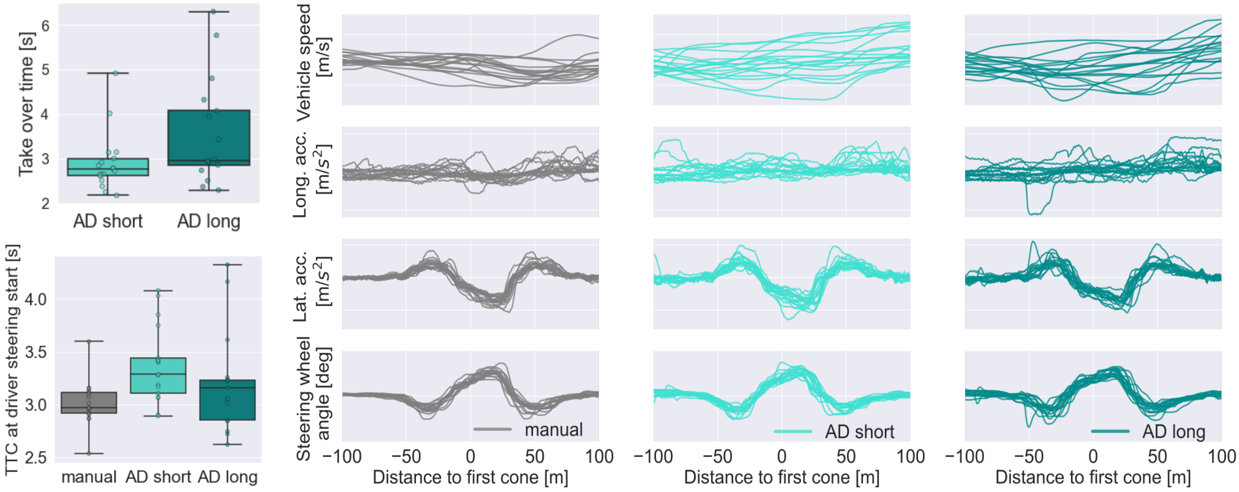

Figure 7.5: Summary of results on take-over response. (a) The take-over time (time from the take-over request until the Traffic Jam Pilot was deactivated) for a short and long automation duration, (b) The time-to-collision (TTC) when the drivers started to steer to pass the first cone in the road-works zone, (c) The driving performance within the road-works zone. Reprinted from "Automation Aftereffects: the influence of automation duration, take-over request and timings", by L. Pipkorn, T. Victor, M. Dozza, & E. Tivesten, 2021b, IEEE Intelligent Transportation Systems. Copyright 2020 by Creative Commons CC-BY-NC-ND 4.0.

7.4.2.2 The driving performance in the road-works zone

All drivers managed to resume manual control in response to the TOR, and then manoeuvre through the cone zone with a similar driving performance as in manual driving without colliding with any cones (see Figure 7.5c). Automation exposure (including both durations) resulted in participants, on average, starting steering to pass the cone zone earlier (at higher TTC) compared to manual (see Figure 7.5b). The effect of automation on the TTC at driver steering start was largest for the short duration, compared to the long duration: the TTC increase was 0.33 s (95% HPD [0.13, 0.55]) for AD short compared to manual, whereas the increase was 0.15 s (95% HPD [-0.063, 0.36]) for AD long and manual.

7.4.3 Discussion and conclusions

The automation aftereffects observed in this study were not as large as previously found in driving simulator studies. For example, previous studies indicate a delayed response after automation compared to manual (e.g. Gold et al., 2013; Louw et al., 2015). In contrast, the present study found that drivers started to steer earlier in response to the road-works zone compared to manual after automation (both durations). To what extent this difference is due to the use of different test environments (driving simulator vs. test track) or different experimental protocols is unknown. However, independent of the test environment, in the search for automation aftereffects it is important to consider the influence of the driver take-over response on the observed aftereffects. That is, more work is needed to disentangle the aftereffects that are merely a result of a longer driver take-over response process, and the aftereffects that may be caused by some other psychophysical human mechanism (e.g. situation awareness, out-of-the-loop, less calibrated sensorimotor control).

7.5 The Effect of the timing of take-over requests on take-over response (L3Pilot Test track study)

The *L3Pilot Test track study was* performed on the ASTA zero rural-road test track using the WoZ vehicle. This section summarizes the study and the findings within the L3Pilot project (see Pipkorn, Tivesten, & Dozza. (2021) for details). This study served as a replication of the previously described ADEST study (see 7.3) with the exception of using L3 automation during the drive that issued a TOR before the lead-vehicle cut-out scenario. The aim of this study was to investigate drivers' response to a lead-vehicle cut-out scenario after a period of L3 automation with a TOR issued at (a) 18 seconds time-to-collision (early) and (b) 9 seconds time-to-collision (late), compared to an (c) adaptive cruise control (ACC; L1) baseline. Further, this study investigated if drivers would crash to the same extent as in the ADEST study (see 7.3) using the same conflict scenario (lead vehicle cut-out and a stationary object).

7.5.1 Methods

The participants drove five laps on the ASTA Zero rural road test track during a drive that lasted approximately 30 minutes. The WoZ vehicle followed a lead vehicle with a 2-second time headway and kept a speed of 70 km/h on straight road segments and down to 50 km/h in curves. The participants were assigned groups with either L3 automation or ACC. The WoZ vehicle simulated L3 automation, and the in-production ACC system was used as a baseline to ensure a consistent time headway in both conditions. The participants were free to engage in any secondary tasks they wanted in the L3 condition and they were instructed to be prepared to take over when requested by the ADF. The participants in the ACC condition were instructed to attend to the driving task at all times. After 30 minutes, the lead vehicle performed a cut-out manoeuvre and revealed a stationary conflict object (a balloon vehicle) to the participants. The participants with L3 automation either received a TOR early at 18 seconds time-to-collision (the early-TOR-condition) or late at 9 seconds time-to-collision (the late-TOR-condition). To avoid a crash the participants had to deactivate automation (early-TOR- and late-TOR-condition) and perform a braking or steering manoeuvre (all three conditions). The number of participants included in the analysis of conflict intervention performance were: N=15 (ACC), N=17 (TOR9) and N=16 (TOR18). The driving performance in the conflict scenario was assessed by (a) response times to the TOR for the drivers' first glance to instrument cluster, first glance forward, hands on wheel, end of secondary task engagement, 2^{nd} try to deactivate automation, automation deactivated, onset of last on-path glance, (b) the steering response time from the conflict object and (c) the vehicle speed and position on road in the interval 200 meters before the conflict object to 100 meter after.

7.5.2 Results

In response to the TOR, participants typically first glanced to the instrument cluster, then either ended their secondary task, looked forward or put hands on the steering wheel, before they deactivated automation. On an aggregated level, the median response time for the first glance to the instrument cluster was 0.68 s (SD = 0.26). Furthermore, the median response time for putting hands on wheel and glance forward was 2.22 s (SD = 1.47) and 2.24 s (SD = 2.50), respectively. Finally, the median response time for automation deactivation was 4.07 s (SD = 2.15). These preparatory actions needed a similar amount of time independent on TOR timing. When the TOR was issued early the drivers typically took longer until their onset of the on-path glance, compared to the late TOR. In addition, when the TOR was early drivers tended to show their first brake before the lead-vehicle cut-out rather than after which was the case when the TOR was issued late.

Three drivers showed extra long take-over times (greater than 10 s). Two participants needed longer time because they did not manage to press the buttons correctly at their first attempt and therefore needed a second try to deactivate automation. One participant was engaged in two secondary tasks simultaneously and was seated with the feet up on the driver's seat. Thus, she took longer time because of the need to change her seating position and put away items before deactivating automation.

All participants successfully managed to avoid crashing with the conflict object. Furthermore, none of the drivers braked to a complete stop. In fact, all drivers passed the conflict object at a speed of about 50-70 km/h. L3 automation with an early TOR resulted in the earliest response, followed by ACC, and L3 automation with a late TOR.

7.5.3 Discussion and conclusions

The results of this study show that drivers resuming manual control after L3 automation may experience a different level of criticality if approaching a conflict scenario at the same time as a TOR is issued, as compared to driving with ACC. The time required for drivers' actions before resuming manual control (e.g. placing hands on the wheel) and the fact that the response process may include off-path glances (e.g. glances to the instrument cluster), might leave very little time for drivers to respond to an upcoming conflict. Since drivers' response process to the TOR includes movement of hands to the steering wheel and off-path glances, it is important that the L3 ADF is capable to take full responsibility for the driving task, including conflict management, until automation is fully deactivated. In addition to deactivating time, all drivers should be provided with enough safety margins to reach the same driving and conflict avoidance performance as in manual driving before required to respond to critical events. Since, drivers are likely not ready to handle any conflict scenario that would occur directly at automation deactivation, the vehicle can potentially provide additional support during this phase e.g. through the use of ADAS.

On the other hand, L3 automation with early TORs may increase alertness and on the overall level generate an earlier conflict response, compared to driving with ACC. The reason for the latter is that drivers coming out of L3 automation through an early TOR are more likely to brake in preparation for the conflict, compared to when a TOR is issued later. Finally, it seems that a TOR during L3 automation clearly communicates the need for drivers to resume manual control and leaves no doubt that the driver is solely responsible for responding to the conflict scenario. The ACC drivers also seem to understand system limitations and the need to act without feedback from the system. This is in contrast to the previous ADEST study with the near-perfect L2 automation system, where many drivers expected that the vehicle would handle the conflict scenario.

7.6 Take-over response on public road (L3Pilot WoZ pilot)

The *L3Pilot WoZ pilot* was performed on the public road (see 7.2.1.2) using the WoZ vehicle (see 7.2.2). The aim of this study was to examine drivers' response process to a TOR in real traffic.

7.6.1 Methods

Thirty participants drove in real traffic with several segments in L3 automation. L3 automation was simulated using the WoZ vehicle. The complete drive consisted of two laps around the Gothenburg ring road (see 7.2.1.2). Each participant experienced automated driving (AD) six times (see Figure 7.6): the first and fourth AD segments lasted 1 minute each and the other four lasted 4 to 6 minutes each. Each AD segment ended with a TOR (TOR 1-6 Figure 7.6). To inform the driver about the need to resume manual control after a period of AD the WoZ vehicle issued a TOR consisting of an audio tone and a visual message in the instrument cluster. The drivers needed to press two buttons on the steering wheel for at least 0.6 s to deactivate automation.

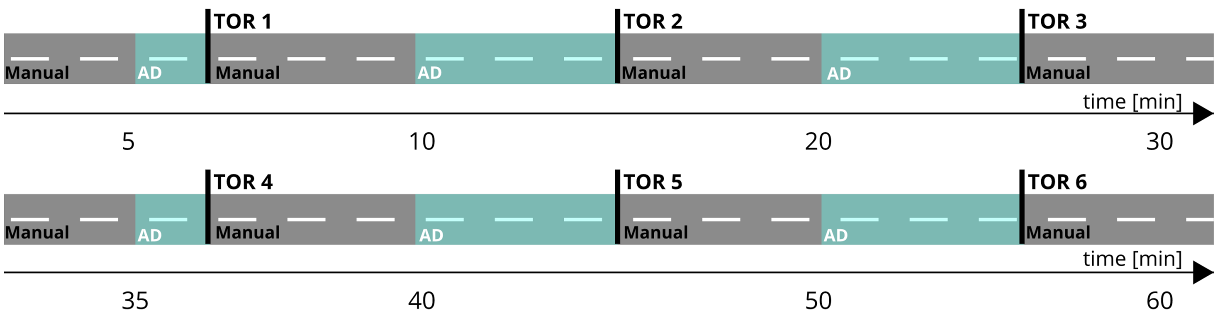

Figure 7.6: The experimental setup for the L3Pilot WoZ pilot

7.6.1.1 The driver response process to the take-over request

To keep the AD duration consistent in the current analysis, only the TORs corresponding to the longer AD segments were used (i.e. TOR 2, 3, 4 and 6 in Figure 7.6). To understand how drivers responded to the TOR, a set of time points for drivers' actions were coded using video data. The coded driver actions were: the first glance to the instrument cluster, the first glance to the forward road, both hands on the steering wheel, automation deactivated and foot on the accelerator pedal. For a case when the driver already looked forward at the TOR, the first glance to forward road was coded as the first glance after an off-road glance (typically to the instrument cluster). The time points for these actions were anchored at the TOR to form response times. The response times aggregated over the four TORs were summarized with median and standard deviation.

7.6.2 Results

7.6.2.1 The driver response process to the take-over request

In 61% of the take-over events the drivers looked towards a secondary task at the time for the TOR. In 34% of the events the driver looked towards the forward road, and in the remaining 5% drivers looked towards any of the mirrors, the instrument cluster or to other off-road glance areas inside or outside the vehicle.

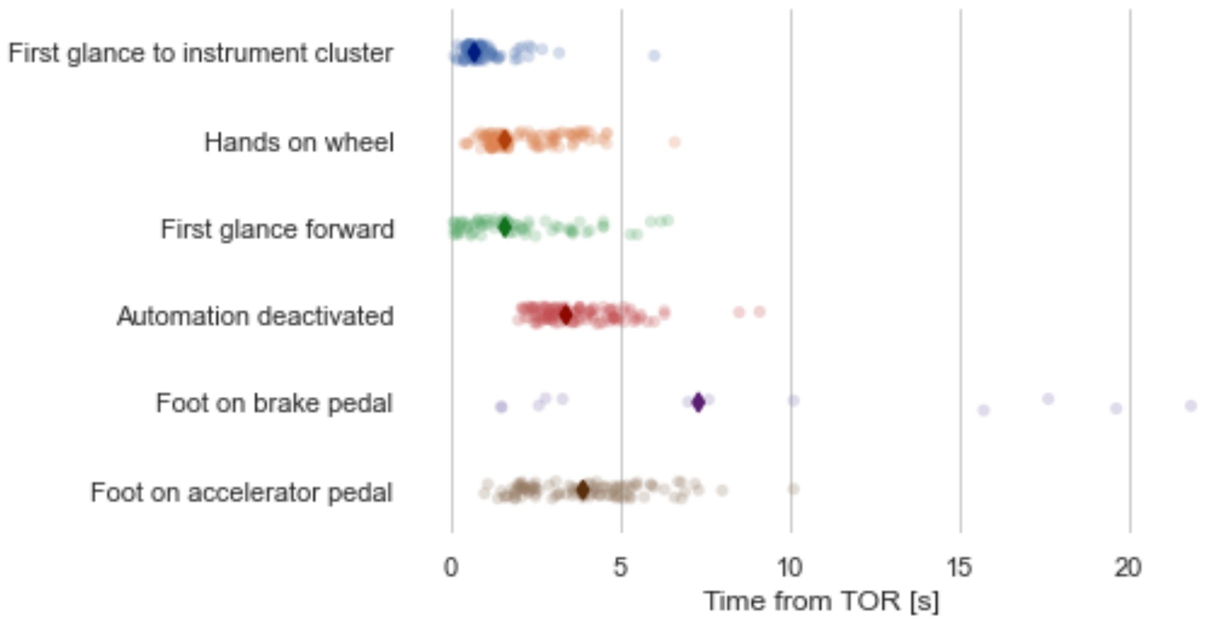

Figure 7.7: The response times to the TOR for first glance to instrument cluster, hands on the steering wheel, first glance to the forward road, automation deactivation, foot on brake pedal and foot on accelerator pedal after a 4 to 6 minute L3 automation duration.

Figure 7.7 shows the response times to a TOR on public road. The median time to show a first glance to the instrument cluster was 0.7 s (SD = 0.72), and the 95th percentile was 2.21 s. In 78% of the take-over events drivers showed their first glance to the instrument cluster before glancing forward and putting the hands on the steering wheel. It was slightly more common for drivers to put hands on the steering wheel before looking to the forward roadway (56%) compared to the opposite order. The median time to put the hands on the steering wheel was 1.6 s (SD = 1.15), and the 95th percentile was 4.12 s. The median time to show the first glance forward was 1.7 s (SD = 1.4), and the 95th percentile was 4.7 s. Further, the median time to deactivate automation was 3.4 s (SD = 1.23), and the 95th percentile was 5.62 s. The longest automation deactivation time (i.e. take-over time) was 9.1s. Whereas all drivers put their foot on the accelerator pedal in response to the TOR, only 10% of the drivers put their foot on the brake pedal within the 30 s after the TOR. The median time to put the foot on the accelerator pedal was 3.9 s (SD = 1.7), and the 95th percentile was 6.8 s.

7.6.3 Discussion and conclusions

In order to deactivate automation in response to a TOR, drivers need to perform certain actions that take some time. In this particular study, some drivers needed up to 9 s before they had resumed manual control. This time partly stems from the physical actions drivers need to perform to be able to deactivate automation, typically, glance to the instrument cluster and place their hands on the steering wheel to press the two buttons. Whereas all drivers eventually did place their feet on the pedals to start to accelerate, almost no drivers were likely to brake in response to the TOR.

7.7 General discussion and conclusions on take-over response

This Chapter presented the findings of the four different studies. The first study (ADEST study) served as a reference since it investigated drivers' conflict response while using a L2 automation system, rather than an L3 automation system. The three other studies investigated drivers' take-over response in three different settings (two on test track and one on public road). Overall, in contrast to the 30% crash rate observed in the ADEST study, no drivers crashed in the TJP study nor in the L3Pilot ASTA study when a TOR was issued prior to the conflict scenario. Thus, our findings suggest that a TOR makes sure the drivers understand that they are in manual control and that they are responsible for handling conflicts when automation has been deactivated.

7.7.1 The process of resuming manual control requires time

Importantly, the process of resuming manual control in response to a TOR during L3 automation take a certain amount of time. Our findings show that drivers, in response to a TOR, typically first glance to the instrument cluster, then either look forward or put hands on the steering wheel before they deactivate automation. The process of a driver responding to a TOR should therefore not be considered as a binary event, but rather as a process with quite large between driver variability. This process to the TOR during L3 automation means that drivers cannot be expected to respond to a conflict until they have completed the take-over response process, deactivated automation, and reached driving performance typical for manual driving. In the L3Pilot Test track study the drivers in the most critical condition were given 6 s for their response preparation process (i.e. look to HMI, look forward, put hands on wheel and deactivate automation). These drivers were all able to start their steering manoeuvre at a similar time as the manual (ACC) group or slightly later. This finding suggests that drivers after automation need at least 6 s for their response process to the TOR to fully resume manual control and be ready to act to events. However, in the same study when the drivers were given 15 s for their preparatory actions, three drivers took 10-12 s to deactivate automation. The reasons behind these extra long take-over times were: (a) a first failed take-over attempt due to a too short button press and (b) engagement in secondary tasks including handheld items. Also, previous research shows that drivers typically take longer time to deactivate automation when given a longer time (Eriksson & Stanton, 2017; Zhang et al., 2019). Therefore, it could be that these drivers would have hastened their response if they had felt a need to do so. Considering the complex response process to TORs during L3 automation, it is important to design L3 automation and TORs that provide drivers with sufficient safety margins to safely deactivate automation and resume manual driving before they are required to respond to potential conflict scenarios.

7.7.2 Response times to TOR on test track and public road

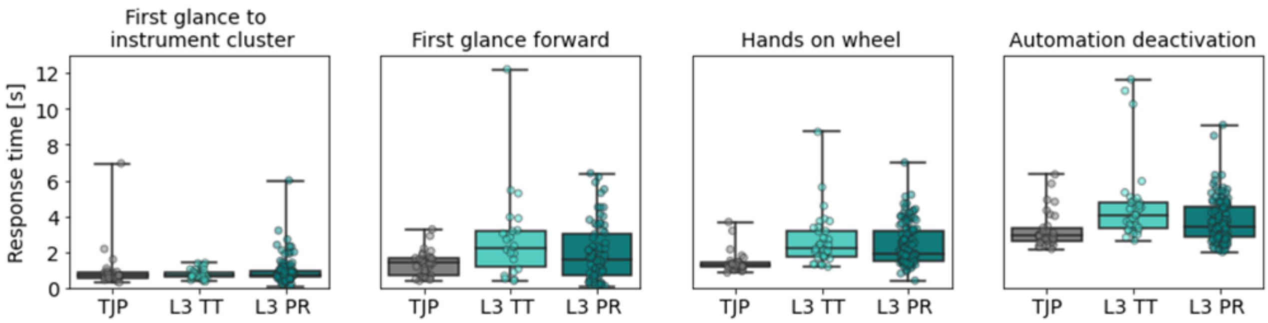

Figure 7.8: Response times to TOR for the three studies: the TJP study (TJP), the L3Pilot Test track study (L3 TT) and the L3Pilot WoZ pilot (L3 PR).

The response times for first glance to instrument cluster, first glance forward, hands on wheel and automation deactivation were measured in the TJP study, the L3Pilot Test track study and the L3Pilot WoZ pilot. Comparison of these response times across studies can give insights into the influence of both environment and TOR design. The TJP study and the L3Pilot Test track study were performed on test track with a conflict scenario present and the L3Pilot WoZ pilot was performed on public road with real traffic. The TJP study included a slightly different TOR design compared to the L3Pilot Test track and WoZ pilot studies since it included seat-belt tensioning as part of the take-over request. The first glance to instrument cluster are similar across studies (see Figure 7.8). The remaining response times were slightly lower for the TJP compared to the L3Pilot Test track and L3Pilot WoZ pilot. This is likely due to the seat-belt tensioning that was used only in the TJP study. It seems that such seat-belt tensioning can hasten the response to a TOR. The response times for hands on wheel and automation deactivation were, however, similar for the L3Pilot Test track and WoZ pilot studies. These findings suggest little influence of the test environment (i.e. test track vs. public road) on the time needed for drivers to put hands on the wheel and deactivate automation. The response time that seemed to differ the most across the three studies were the time needed for drivers to glance forward. The quicker response observed for drivers in the L3Pilot WoZ pilot compared to L3Pilot Test track study could be due to the more dense traffic on public road compare to test track.

7.7.3 Limitations and recommendations

In the three studies including L3 automation all drivers responded to the take-over request and deactivated automation. Thus, the response times presented here should be viewed as response times for systems when drivers successfully respond to the take-over request. However, the critical question of what happens if the user voluntary or involuntary fails to respond to a take-over request remains unknown. Using a WoZ setup makes sure no critical situations occur in case of an unresponsive user. The WoZ driver would still be able to control the driving task and consequently serves as a safety back-up. Future L3 automation systems should preferably be able to detect if a driver is not fallback-ready and activate a safety back-up response to prevent the risk of a crash.

8 Driver impairment study

Chapter 8 presents results from analyses of uneventful driving in two different test track studies. The studies were planned by Chalmers and VCC, and the data was collected by VCC staff. The participants were VCC employees that neither work as professional drivers nor with development of vehicle automation.

8.1 Aim and research questions

This chapter focuses on the assumption of a fallback-ready user in L3 automation and the impact which certain driver states, normally viewed as severely performance-degrading, might have on that assumption. For example, while sleepiness generally increases as a function of manual driving time, it remains to be determined whether it will be harder or easier for drivers to stay alert (and consequently fall-back ready) when using L3 automation. Drivers can disengage from the driving task in L3 automation (possibly leading to underload) but are free to engage in and select other tasks while the automation is enabled (possibly counterbalancing underload).

There is also an additional measurement problem in relation to driver impairment, when using L3 automation. In manual driving, severe sleepiness and high levels of Blood Alcohol Concentration (BAC) will influence the driver's lane keeping performance (Pilutti & Ulsoy,1999; Lee et al., 2010), and this performance degradation can be captured using in-vehicle sensors measuring the distance to the left and right lane markers. In L3 automation, the ADF performs all parts of the driving task as long as the automation is activated and tracking of lane markers thus says nothing about the occupant's status. Consequently, other means of sensing are needed to recognize signs of severe driver impairment. These could include detecting breath alcohol in the cabin or tracking specific patterns in drivers' visual behaviour by means of a driver monitoring system (DMS).

Within this rather wide scope, four specific research questions are addressed within this chapter:

RQ-U5 What is drivers' level of fatigue while using the ADF?

More specifically, this chapter analyses the effect of automation (SAE L0, L1, L2, L3) on driver sleepiness as a function of drive time. In the L3pilot ASTA study, L3 is compared to L1 during a drive. In the intoxication study L3 is compared to L0, and L2.

RQ-U5E1 What is the effect of high alcohol intake on driver sleepiness as a function of drive time in different levels of automation?

RQ-U6 What is the effect of ADF use on driver attention to the road/other road users?

In the intoxication study L3 is compared to L0, and L2 during segments with and without instructed secondary tasks.

RQ-U6E1 What is the effect of high alcohol intake on driver's visual behaviour in different levels of automation?

In the intoxication study a first sober baseline drive is compared to a second intoxicated drive (BAC 0.1%) for three groups assigned to L0, L2, or L3 automation including segments with and without instructed secondary tasks.

Section 8.3.1 presents descriptive statistics on self-reported sleepiness addressing RQ-U5 and RQ-U5E1. Section 8.3.2 presents the general findings and conclusions on visual behaviour (RQ-U6/RQ-U6E1), while more detailed results can be found in Tivesten, Broo, and Ljung Aust (2021).

8.2 Methods

8.2.1 Data collection

Table 8.1 provides an overview of the studies and how they link to the research questions formulated within the L3pilot project. Both test track studies included in this chapter were performed on ASTA zero rural road test track described in 7.2.1.

Table 8.1: Details about the two studies presented in this chapter.

	L3Pilot ASTA study	Intoxication study
Test environment	Test track	Test track
Test conditions (during uneventful driving)	Speed ≤ 70 km/h Lead vehicle present No instructed secondary tasks Sleepiness reported every 5 min	Speed = 50 km/h No other vehicles present Three instructed secondary tasks during each drive Sleepiness reported every 7 min
Levels of automation	• L1 (ACC) • L3 (simulated ADF)	• L0 (manual) • L2 (Pilot Assist, PA) • L3 (simulated ADF)
Driver conditions, and test duration(s) for each participant	• One drive: Normal, 30 min	• 1st drive: Normal, 30 min Session with alcohol intake • 2nd drive: BAC ≈ 0.1%, up to 60 min
Karolinska sleepiness scale (KSS)	At start of the drive and then every 5 min	At start of the drive and then every 7 min
RQs of interest	RQ-U5 (fatigue)	RQ-U5 (fatigue) & U6 (attention)

Note that the self-reported sleepiness was obtained for all participants that completed the drive, while the analysis relying on video or vehicle signals were not available for all participants (e.g., missing data, driver wearing sunglasses). Therefore, the number of participants reported in the following sections may vary dependent on the analysis.

The L3pilot ASTA study

The method used in the L3pilot ASTA study are described in more detailed in chapter 7, which is focused on driver take-over and conflict response process (see section 7.5.1). The analysis in this chapter focuses the uneventful driving taking place before the take-over request analysed in chapter 7. In addition to the method description in chapter 7, some additional details are presented here that are relevant for the research questions addressed in this chapter. ACC served as a baseline to obtain the same car following distance relevant to the research questions addressed in chapter 7 (i.e., take over and conflict avoidance performance).

All participants completed a 30-minute drive on test track following a lead vehicle at speeds up to 70 km/h. The participants either used ACC (N=20), or L3 automation (N=38) throughout the drive. The participants in L3 were free to engage in secondary tasks of their own choice, while the participants in ACC were instructed to attend to the driving task. None of the drivers were instructed to perform any specific secondary tasks.

The participants reported their subjective level of sleepiness during the drive using the Karolinska Sleepiness Scale (KSS; Åkerstedt and Gillberg, 1990). KSS is a 9-point scale ranging from 1 (extremely alert) to 9 (extremely sleepy – fighting sleep). The participants reported KSS at the end of each lap (every 5 minutes, labelled as Lap 1 – Lap 5) and at the very beginning of the drive (labelled as Start).

The Intoxication study

The test vehicle was a Volvo XC90 equipped with additional sensors, loggers, and a double command on the passenger side. A vehicle speed cap was implemented to restrict the maximum speed to 50 km/h (for safety reasons). The SAE level 2 (L2) driving mode used the in-production Pilot Assist function and settings. Pilot Assist supports the driver in lateral and longitudinal control, but the drivers need to keep their hands on the wheel and is always fully responsible for the driving task. A hands on wheel reminder is issued in case there is no driver torque input detected. The SAE level 3 (L3) was simulated using the PA function, but with the hands on wheel reminder removed. The participants were also instructed that they could disengage from the driving task when the L3 mode was on, but they needed to be prepared to take-over control if requested by the ADF.

There were no other vehicles present on the test track during the intoxication study. Each participant was assigned to either manual driving (L0, N=11), Pilot Assist (L2, N=11), or L3 automation (L3, N=10).

Each participant first performed a sober baseline drive by completing 4 laps on the test track that lasted for approximately 30 minutes. All participants drove manually the first 5 minutes, and then activated automation (L2 or L3) or continued driving manually according to their assigned group. The participants were instructed to perform three different secondary tasks (manual radio tuning, calling a phone number, adjusting the set temperature on the driver side) using the centre stack display on straight road segments between 6 to 15 minutes into the drive, while the last 15 minutes of the drive continued without instructed tasks.

After completing the baseline, participants drank alcohol for 45 minutes and then waited another 15 minutes to reach the BAC target of 0.1%. The following intoxicated drive replicated the baseline drive during the first 30 minutes and then continued for another 35 min of uneventful driving unless the participant needed to take a restroom break or end the drive for other reasons (e.g., if feeling severely drowsy, or sick). All participants completed the first 4 laps without breaks. The participants reported their level of sleepiness (KSS) after each completed lap (every 7 minutes) and at the beginning of each drive. A calibrated Breath Analyzer was used to estimate blood alcohol concentration by measuring breath alcohol upon arrival at the test track, just before starting the second drive, and just after completing the second drive.

8.2.2 Analysis of driver sleepiness

The mean and standard deviation of KSS were plotted for the reported instances during the drive. In addition, histograms were plotted for each condition to show the distribution of individual change in KSS from the start of the drive to the end of the last lap included in each analysis. KSS were analysed separately in the two studies due to some differences in the experimental design (see table 8-1 for overview) and should be considered as descriptive statistics, since no corrections for multiple tests were performed.

The L3pilot ASTA study:

The reported change in KSS from start to the end of the drive was compared between ACC-mode (ACC, N=20) and L3 automation (L3, N=38) using a t-test.

The intoxication study:

A paired t-test was used to compare the change in KSS from start to end of the drive in the baseline drive compared to the intoxicated drive. A one-way ANOVA was used to compare KSS change during the drive for the three levels of automation (L0, L2, L3) in the sober and intoxicated drive separately.

8.2.3 Analysis of visual attention (Intoxication study)

To address the specific research questions formulated in section 8.1, five segments per drive were selected for manual annotation of the drivers' glance behaviour. The start of segments S2-S4 was defined by the onset of the first off-path glance towards the secondary task, and the end of the segments was defined by the end of the last glance towards the task. Segments S1 and S5, which did not contain any instructed secondary tasks, were selected from the same stretch of road at the beginning and at the end of each drive.

Table 8.2: Overview of selected segments during the first sober baseline (BL) drive, and the second drive while intoxicated (IN).

Drive	Level of automation	Segment	Instructed tasks	Duration	Lap
BL	L0	S1	No	30 s	1
	L0, L2, or L3	S2	Radio	12 - 50 s	2
		S3	Dial	13 - 30 s	2
		S4	Temp	5 - 22 s	2-3
		S5	No	15 s	4 (last lap)
IN	L0	S1	No	30 s	1
	L0, L2, or L3	S2	Radio	12 - 50 s	2
		S3	Dial	13 - 30 s	2-3
		S4	Temp	5 - 22 s	2-4
		S5	No	15 s	7-9 (last lap)

Drivers' glance locations were first coded as a timeseries and then transformed to a binary eyes on/off path signal. Instances where the gaze location could not be determined (i.e., when the eyes were not visible) were treated as missing data. The effects of intoxication and automation on the participants visual behaviour was then analysed using descriptive statistics and non-parametric statistical tests. This deliverable reports on the main conclusions from the study. For a detailed method description and results, see Tivesten et al. (2021).

The glance metrics considered for all segments were the following:

- Percent road centre, PRC [%]: Percent of time with eyes on path (e.g. on forward roadway).
- Off-road glance Frequency, GF [N]: Number of off-path glances.

The glance metrics considered for all task segments (S2-S4) were the following:

- Total glance time, TGT [s]: The sum of all off-path glance durations
- %GD>2s [%]: Percentage of off-path glances longer than 2s.
- MaxGD [s]: Maximum off-road glance duration.

8.3 Results

8.3.1 The influence of automation and alcohol intake on drivers sleepiness (RQ-U5/ RQ-U5E1)

This section includes an analysis of how KSS changes over time during a 30-minute drive, based on the L3pilot ASTA study and the intoxication study.

The L3pilot ASTA study:

Figure 8.1 shows that the average KSS were similar at the very beginning of the drive (labelled as Start) for both groups of participants, using either L3 automation or ACC that served as a baseline in this study. KSS ratings increased slightly during the 30-minute drive for both groups. Participants driving with ACC reported slightly higher KSS values at the end of the drive (N=20, M = 5.05, SD = 1.47) compared to drivers using L3 automation (N = 38, M = 4.58, SD = 1.48) labelled as Lap 5 in Figure 8.1.

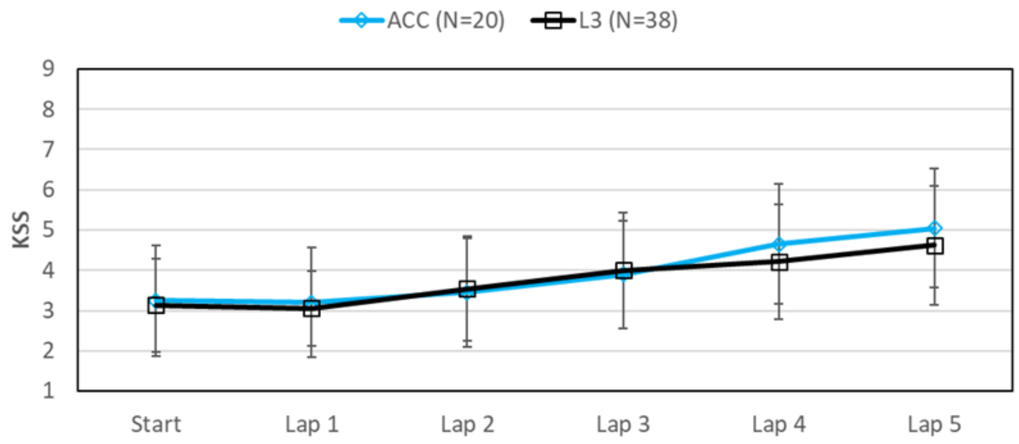

Figure 8.1: KSS ratings, reported at the very start of the drive (Start) and then at the end of each lap on the test track (Lap 1 - Lap 5) in the L3pilot ASTA study for participants using ACC or L3 automation. The markers show mean values and error bars shows the standard deviation.

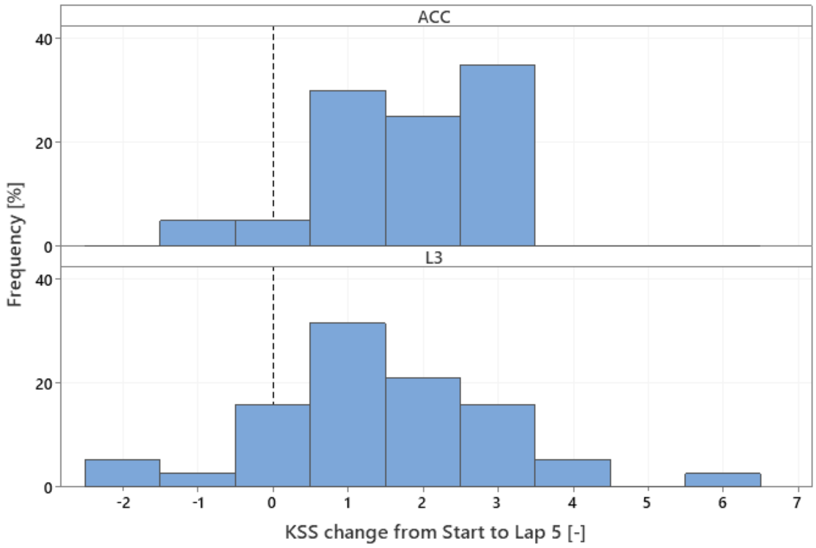

Figure 8.2: Histogram of KSS change from start to end of the drive for participants using adaptive cruise control, (ACC) or L3 automation (L3) through the complete drive.

Consequently, the average change in KSS from the start to the end of the drive was slightly lower in L3 automation (M = 1.45, SD = 1.61) than in ACC (M = 1.80, SD = 1.15) though this difference was not statistically significant ($t(50) = 0.96$, $p = 0.34$). A slightly larger standard deviation was also observed in L3 automation compared to ACC as illustrated in Figure 8.3. In other words, while some drivers using L3 got a little more tired than those using ACC, other L3 drivers actually got more alert.

The intoxication study:

The reported KSS was slightly higher at the start of the intoxicated drive (second drive) compared to the start of the baseline drive (first drive), and the average KSS had a similar increase during both drives as illustrated in Figure 8.3.

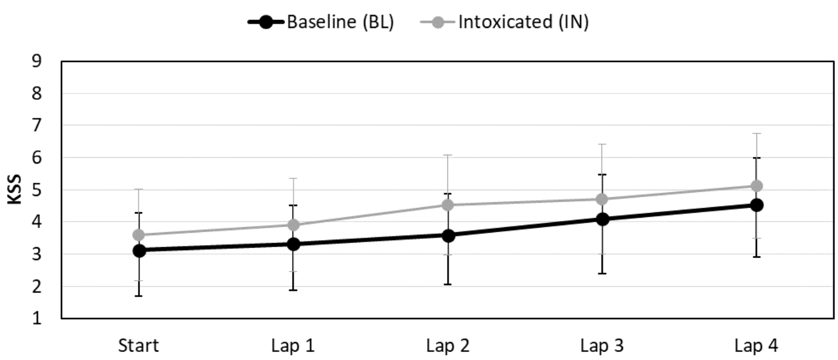

Figure 8.3: KSS rating reported at the start of the drive and at the end of each lap on the test track (Lap 1-4) for the first sober drive (baseline) and the second drive at approx. BAC 0.1 (intoxicated)

when all levels of automation (L0, L2, L3) were combined. The markers show mean values and error bars the standard deviation.

The participants' average change in KSS from the start to the end of the drive, was slightly lower in the baseline (N=32, M = 1.34, SD = 1.12) than in the intoxicated drive (N = 32, M = 1.53, SD = 1.78). However, this difference was not statistically significant (t (31) = -0.61, p = 0.55).

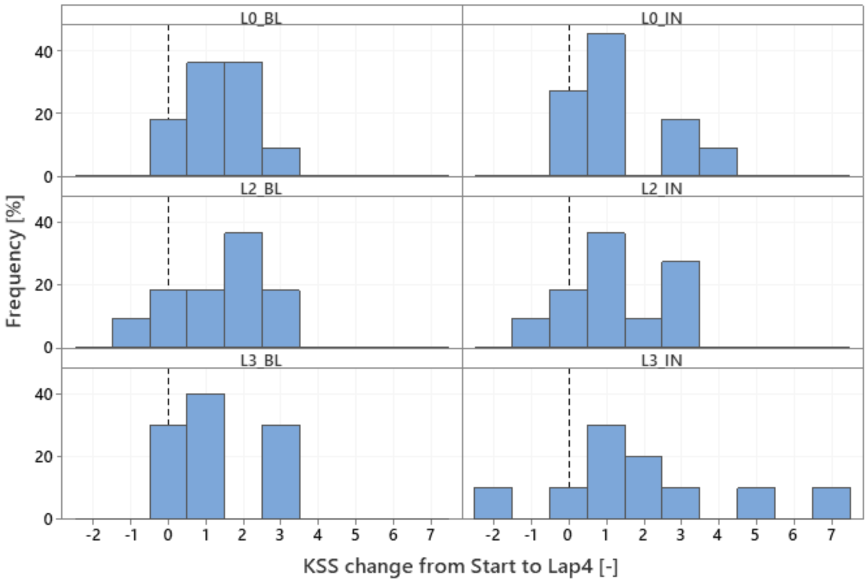

Figure 8.4: Histogram of change in KSS rating from start of the drive to end of the fourth lap on test track panelled by baseline and intoxicated drive (BL, IN) and level of automation (L0, L2, L3).

The KSS change was similar for all levels of automation, but the mean and standard deviation was slightly higher in the L3 intoxicated condition compared to the L3 baseline and all drives in lower levels of automation (see Figure 8.4). However, there were no statistically significant differences in KSS change due to level of automation during the sober drive (F (2,29) = 0.01, p = 0.99) nor during the intoxicated drive (F (2,29) =0.50, p=0.61).

8.3.2 The influence of automation and intoxication on visual behaviour (RQ-U6)

When intoxicated, drivers showed higher PRC values when not doing instructed secondary tasks (segments S1 and S5) and lower PRC values when doing them (S2-S4) as compared to when driving sober (see Figure 8.5). Note that all participants drove manually in segment S1 at the beginning of the drive, while they drove according to their assigned group (L0, L2, or L3) during the remaining segments including S5 at the end of each drive. The off road glance frequency (GF) was generally lower in all segments during the intoxicated drive compared to the baseline drive.

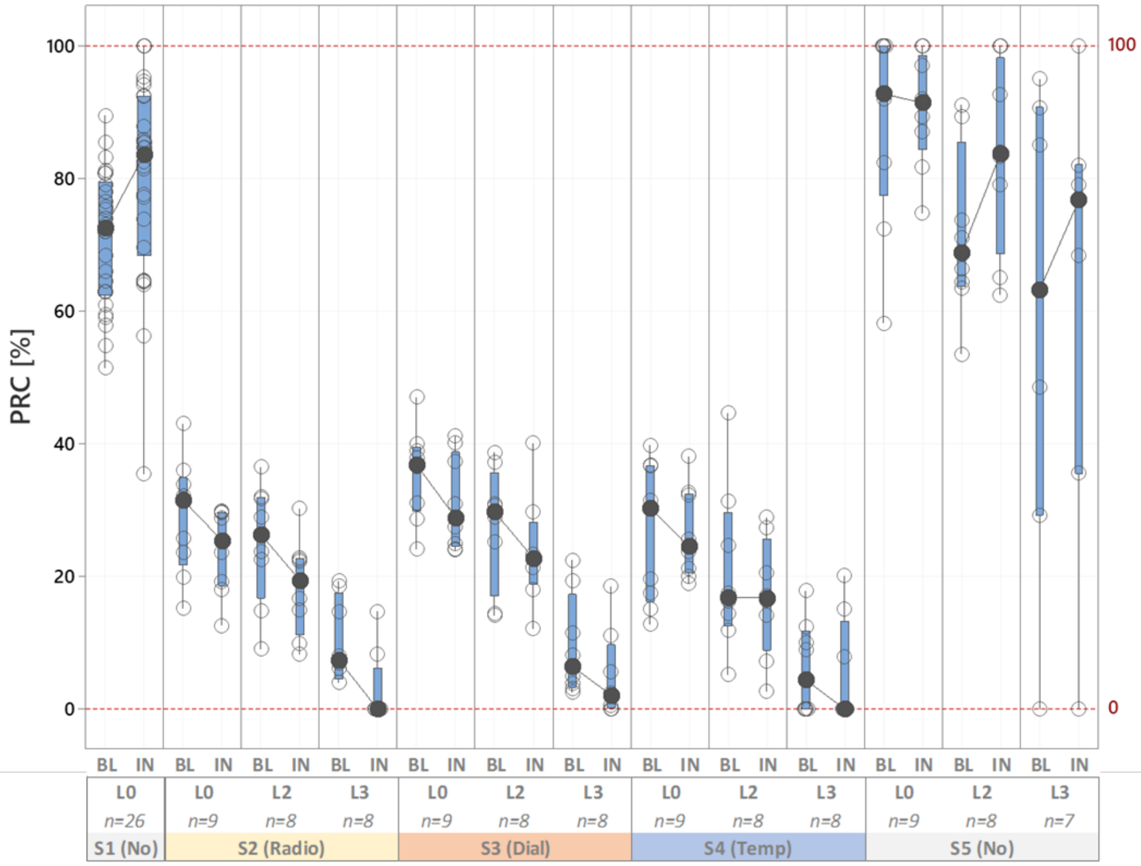

Figure 8.5: Boxplots of PRC including individual and median markers. Each panel shows the baseline (BL) and the intoxicated drive (IN) where the medians are connected with lines. The panels are divided by level of automation (L0, L2, L3) and segment (S1-S5).

The off road glance durations increased with both level of automation and intoxication. Figure 8.6 shows the cumulative distributions of off path glance durations for the three secondary tasks during both the basline and the intoxicated drive.

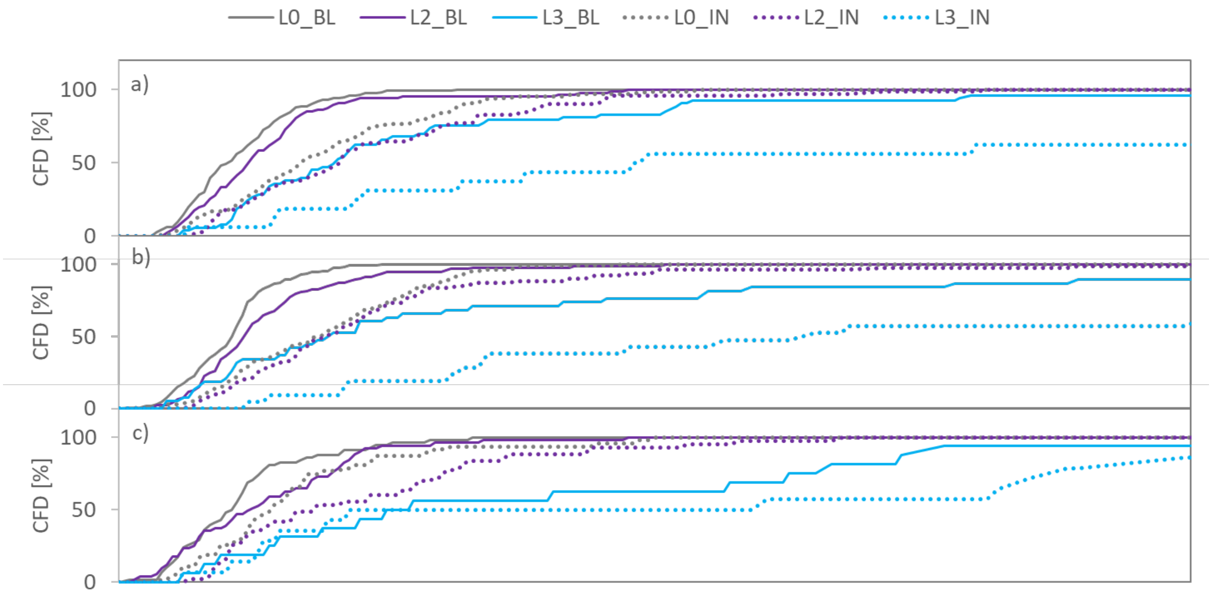

Figure 8.6: Cumulative frequency distribution (CFD) for off path glance durations during the a) radio task, b) dialling task, and the c) temperature task. Solid lines represent the sober baseline drive (BL), while dotted lines represents the intoxicated drive (IN). The colours represent the different levels of automation including L0 (grey), L2 (purple), and L3 automation (light blue).

The radio and the dialling task turned out to be more visually demanding than the temperature task, since the median TGT and GF were about twice as high for the first two tasks compared to the temperature task during the baseline drive in manual mode.

The effect of intoxication on drivers' glance behaviour was most evident during the more visually demanding secondary tasks and mainly influenced the long glance metrics (MaxGD, %GD>2s) and to some extent the TGT.

The effect of automation on the drivers' glance behaviour was present in all segments and most evident for the secondary task segments. As expected, the difference in glance metrics was most apparent when comparing L3 with L0, while the difference between L0-L2 was much smaller compared to the difference between L2-L3. The glance metrics PRC, GF, MaxGD, and %GD>2s were all influenced by level of automation in both the baseline and the intoxicated drive.

8.4 Discussion of results and conclusions

On average, sleepiness seems to increase at a similar rate as a function of drive time in different levels of automation, as well as when driving with or without alcohol intoxication. However, there was an observed trend that a few individuals had larger changes in KSS, both in positive and negative direction, as a consequence of higher levels of automation or alcohol intoxication. If these results are corroborated in further studies, i.e. that L3 automation could result in larger variations of how sleepiness develops as a function of drive time, it could potentially mean that more drivers are at risk of falling asleep at the wheel while using L3, as compared to manual driving. Further studies of how severe sleepiness develops in real traffic, how it can be detected, and what countermeasures are effective in maintaining the driver in the fall-back ready state required for L3 usage, are thus needed.

The drivers’ glance behaviour was substantially affected by the level of automation, producing lower PRC in all segments and longer off path glances during secondary tasks in L3 compared to L0 and L2. Note though that the change in glance behaviour between L2-L3 was much larger than the changes between L0-L2. Alcohol intoxication seemed to influence the glance metrics in the same direction, further amplifying the effects seen during the task segments when using automation. On the other hand the drivers tended to look more on the road during non-task segments when they were intoxicated. These findings suggest that glance metrics based on an eyes on/off road signal may be one out of several indicators that could potentially detect alcohol intoxication in different levels of automation. More advanced metrics derived from driver monitoring sensors (DMS), such as gaze dispersion/concentration or nystagmus (i.e., involuntary jerky eye movement), may also prove sensitive to detecting severe alcohol intoxication. However, a DMS requires drivers to mainly look forward to correctly capture drivers’ eye movements. These systems may then be limited in detecting severe alcohol intoxication if drivers decide to look away from the road during almost the complete automation duration, as found for some drivers in the L3pilot WoZ pilot study (Pipkorn, Dozza & Tivesten, 2021).

9 Driving Simulator Study on short-term behavioral adaptation

In this chapter a driving simulator study is described that explored the change of manual driving behaviour directly after driving with an L3-ADF. The study was conducted by Leeds University.

9.1 Aim and research questions

The main aim of this driving simulator study was to understand whether experiencing automated car-following influences drivers' subsequent manual car-following behaviour. The two predominant factors contributing to rear-end collisions are a driver's failure to perceive and/or react to a lead vehicle's action, likely to be exacerbated by close car-following behaviour (Dingus et al., 2006). However, these two factors have not yet been systematically investigated in the context of behavioural adaptation (BA) and vehicle automation. Therefore, to address the first aim of this study, an urban car-following scenario was created, where all drivers were exposed to one of two time headway (THW) conditions (0.5 s vs 1.5 s) maintained by a highly automated vehicle. We assessed whether exposure to these two THWs changed drivers' adopted THW in a subsequent manual car-following situation, compared to their initial THW in manual driving, before automation was experienced. These THW parameters were based on the 25th and 75th percentile of a driver behaviour model, based on naturalistic driving studies, which incorporate drivers' instantaneous aggressiveness during car-following scenarios (Niels, Edoardo, Florent, & Clément, 2019). Our aim was to expose drivers to two fairly 'aggressive' automated car-following scenarios. A 1.5 s THW has been used in other studies (cf. de Waard et al., 1999; Lyu et al., 2018; Heikoop, de Winter, van Arem, & Stanton, 2019). We avoided longer THWs, to ensure that drivers did not feel too disconnected from the lead vehicle. The shorter 0.5 s THW was chosen to allow an observable comparison in behaviour with this headway. We hypothesised that, overall, drivers will reduce their THW in manual car-following after experiencing automated car-following, but that this reduction will be greater after experiencing the shorter THW.

The second aim of this study was to understand how engagement with the driving task during automated car-following influenced whether drivers changed their THW in subsequent manual car-following. We hypothesised that drivers in L2 automation, who are expected to continuously monitor the road environment, would be more susceptible to changing their THW after automated car-following, than drivers in L3, who were encouraged to look away from the road environment, and were perhaps not aware of the two automated headways.

Given the emphasis on personal characteristics in determining susceptibility to BA, and driving style, more generally (Itkonen & Lehtonen, 2020), we also investigated whether changes in THW would co-vary with drivers' self-reported traits, including sensation seeking (Arnett, 1994), traffic locus of control (Özkan & Lajunen, 2005), and driver style questionnaire (French, West, Elander, & Wilding, 1993). Drivers with an external LOC and who scored high on the SS scale were hypothesised to be more likely to exhibit BA. Our primary research questions were:

1. Do drivers change their car-following behaviour in manual driving after experiencing car-following in automated driving?

2. Is this influenced by the THW adopted by the automated driving system?
3. Is this influenced by whether drivers resume control in the presence of a lead vehicle?
4. Is this influenced by engaging in a visual NDRA during automation?

9.2 Methods

9.2.1 Participants

Following approval from the University of Leeds Research Ethics Committee (Reference Number: LTTRAN-054), we recruited two groups of 16 drivers, via the driving simulator database. Participant details for each group are displayed in Table 9.1. Participants received £25 for taking part in the experiment and were free to withdraw at any point. Three participants were not considered for analysis, as they did not adhere to the experiment instructions to follow the lead vehicle. One participant was excluded because of missing data. This leads to N=28 valid participants.

Table 9.1: Participant demographics information.

Demographics	Gender Mean (SD)		Automation Group, Mean (SD)	
	Males (N=19)	Females (N=9)	L2 (N=15)	L3 (N=13)
Age (years)	39 (16)	38 (10.83)	42 (17)	33 (8)
Miles travelled annually	11368 (9401)	7763 (4302)	8753 (4719)	9116 (8200)
Years of driving experience	19 (15)	16 (8)	22 (16)	14 (8)

9.2.2 Design and Procedure

9.2.2.1 Equipment

The experiment was conducted in the full motion-based University of Leeds Driving Simulator (UoLDS), which consists of a Jaguar S-type cab housed in a 4m diameter spherical projection dome with a 300° field-of-view projection system. The simulator also incorporates an 8 degree-of-freedom electrical motion system. This consists of a 500mm stroke-length hexapod motion platform, carrying the 2.5T payload of the dome and vehicle cab combination, and allowing movement in all six orthogonal degrees-of-freedom of the Cartesian inertial frame. Additionally, the platform is mounted on a railed gantry that allows a further 5m of effective travel in surge and sway.

When active, the ADF assumed lateral and longitudinal vehicle control and maintained a maximum velocity of 40 mph. However, in the presence of a slower lead vehicle, the system would reduce its speed, to maintain the time headway of the respective condition (described below). The status of the ADF was indicated by the colour of a steering wheel symbol that was located on the left panel of the central display unit (Figure 9.1). During the automated drives, the steering wheel symbol was solid green when automation was engaged, and red when automation was unavailable.

Figure 9.1: An example of the in-vehicle HMI with the automation status symbol (Left: Automation not engaged, Right: Automation engaged) and the vehicle speed (mph).

9.2.2.2 Experimental Design

A 2X2X2 mixed design was used for this study, with a between-participants factor of *Level of Automation* (L2, L3) and within-participant factors of *Time headway* (Short: 0.5 s, and Long: 1.5 s) and *Take-over type* (with lead car, without lead car). All factors were fully counterbalanced.

Level of Automation determined the activities drivers were permitted to do during automated driving. Participants in the L3 group were instructed to engage in a visual non-driving related "Arrows" task (NDRA) during automation (Jamson & Merat, 2005). The Arrows task required participants to search for, and touch, the upward-facing Arrow, displayed in a 4x4 grid of Arrows, using a touch screen in the centre console. The screen displayed the current participant's cumulative score and a 'score to beat' to keep them engaged in the task. Participants were also told they would get an additional £5 if they beat the best score, though, for ethical reasons, all participants received this reward at the end of the experiment, regardless of performance. The Arrows task was only available when automation was engaged.

Take-over type specifies whether drivers resumed control during a car-following, or free-following, scenario. For all experimental drives, approximately two minutes after drivers engaged automation, a lead vehicle moved into the ego vehicle's lane, from an adjacent road, triggering automated car-following. However, for half of the trials, the lead vehicle continued in its path when the transition to manual control was triggered, while for the other half of the trials, the lead vehicle exited the lane a few moments before the take-over event (see Figure 9.2). For the trials without a lead vehicle, a new lead vehicle joined the ego vehicle's lane, from an adjacent road at the next intersection, which was 20 m from the previous intersection. The aim of this manipulation was to assess whether, after the resumption of control, drivers would attempt to catch up, and then maintain the same headway with a new lead vehicle, as the headway assumed with a vehicle immediately ahead of them. Note that a late resumption of control never led to a crash, unless drivers sped up after resumption of control, since the lead vehicle always assumed a safe headway.

9.2.2.3 Procedure

Upon arrival, participants were briefed on the description of the study and were asked to sign a consent form, with an opportunity to ask any questions. They were then given a chance to practice manual driving, and automated driving, within a 2-lane urban road, with low-density oncoming traffic. During the practice session, participants were talked through the various aspects of the vehicle HMI, were shown how to engage and disengage the automation and, those in the L3 condition practiced the Arrows task.

Participants were asked to drive in the centre of the lane and maintain the 40 mph speed limit. They were asked not to overtake any lead vehicles, but to otherwise adhere to the standard rules of the road, ensuring safe operation of the vehicle, and maintaining their desired distance to the vehicle ahead. Before the start of the automated drives, participants were presented with an auditory-verbal request to engage automation: "*Attention engage automation*". To engage the ADF, participants pressed a button on the steering wheel, after which they took their hands away from the steering wheel and foot away from the accelerator. At the end of the automated drives, participants were presented with an auditory-verbal take-over request, "*Attention, get ready to take-over*". The TOR was presented when the vehicle reached a section of road with faded road markings, which represented a system limitation condition, and a need to resume control. After this alert, a short duration acoustic tone (1000 Hz, lasting 0.2 s) sounded with increasing frequency until participants resumed manual control. Participants could disengage automation by either pulling the stalk, moving the steering wheel (threshold of 2° was applied), or pressing the brake, or accelerator pedals. Our aim was to implement a non-critical take-over request that did not cause drivers any distress. The road markings reappeared shortly after drivers resumed control. All drivers resumed control, and the exact take-over time varied according to when drivers resumed control, but it was generally between 10-13 s.

Following the practice drive, participants completed two experimental runs (see Figure 9.2). Run 1 consisted of five different, but connected, driving segments, starting with a brief Manual Baseline Drive (~6 min) which started with a ~4-minute free-driving scenario, and a ~5-minute car-following scenario, after which the lead vehicle turned off the road and drivers carried on driving for ~1-minute. This period was used to collect 'baseline' data for drivers' THW during car-following and was only included in Run 1.

Apart from Manual Baseline Drive, the sequence of events for Run 1 and Run 2 were identical. Each driver experienced the following order of events: Automated Drive 1 (~ 5 min), Manual Drive 1 (~5 min), Automated Drive 2 (~5 min), and Manual Drive 2 (~5 min). Run 2 began with a brief period of manual driving to allow participants to engage the ADF. Experimental run 1 and 2 were counterbalanced, which varied the order in which drivers experienced long and short THW automated car-following. Within each drive, whether or not drivers resumed control in the presence of a lead vehicle was also counterbalanced.

To reduce the effect of fatigue, a short break was introduced after the practice drive and experimental drives. After each of these drives, participants were taken out of the driving simulator, and asked to complete a three-part questionnaire, which included the Arnett Inventory of Sensation Seeking (AISS; Arnett, 1994), traffic locus of control (T-LOC; Özkan & Lajunen, 2005), and driver style questionnaires (DSQ; French, West, Elander, & Wilding, 1993). Finally, after Run 1 and Run 2, respectively, drivers rated their perceptions of their own and the ADF behaviour during the preceding drive (either Long THW or Short THW) by indicating on a five-point Likert scale (1: “Strongly disagree” to 5: “Strongly Agree”) their level of agreement with the following statements,

1. During the automated drive, the system kept a safe distance from the car in front;
2. During the automation drive, I think the system should have kept a closer distance from the car in front;
3. During the automated drive, I think the system should have kept a long distance from the car in front;
4. Experiencing the automated driving system changed how I drove in the subsequent manual drive;
5. Following the automated drive, when there was a vehicle in front of me, I used the accelerator and brakes more than normal;
6. I kept the same distance to the vehicle in front during the manual drive as I experienced in the automated drive.

The entire experiment lasted approximately 2.5 hours.

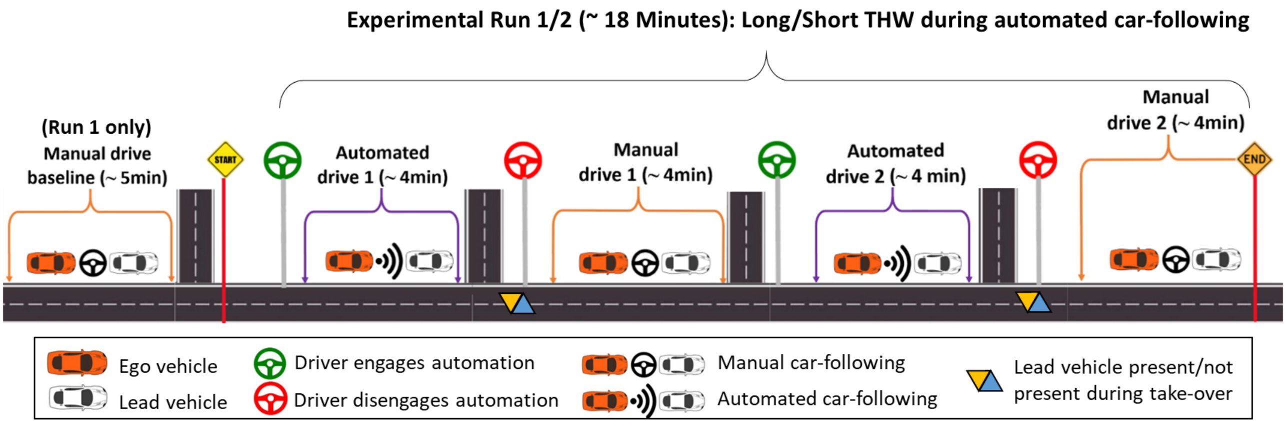

Figure 9.2: Schematic representation of the two experimental runs, which exposed drivers to automated car-following with either a long (1.5 s) or short (0.5 s) time headway. Each run comprised of two sequences of automated and manual car-following drives. Only the first run of the experiment included a manual baseline car-following drive. Between each drive, drivers had to

take-over control, either with or without a lead vehicle. The order of the runs and presence of the lead vehicle during the take-over, was counterbalanced across participants.

9.3 Analysis

9.3.1 Establishing car-following

In order to analyse drivers' car-following behaviour, following resumption of manual control from automation, we first needed to establish that they had stabilised their control of the vehicle, and were engaged in a consistent car-following behaviour. The concept of stability in car-following was initially proposed by Herman et al. (1959) and is characterised as a consistent variation in drivers' following distance, which does not affect the overall microstructure of the surrounding traffic. We calculated the point at which drivers had entered a stable car following period, labelled "stabilisation time", using an algorithm developed by Gonçalves et al. (2020). In this work, the metric was measured as the time between the take-over, and the point at which drivers' average THW remained below a particular threshold, for at least 10 s. This threshold was based on inflexion points in the overall distribution of the THW during the whole car-following task, for each driver. We used this technique to calculate stabilisation time for both take-over types, i.e., irrespective of whether or not there was a lead vehicle during the take-over.

To establish car-following events for our analysis, we considered driving data from the stabilisation time to the moment the lead vehicle left the road. According to Gipps (1981), a car-following task is characterised by a constant mediation, and adjustment, of drivers' distance to the lead vehicle, according to their desired safety boundaries and willingness to increase their speed. Therefore, we filtered out the sections of manual driving when drivers were too far away from the lead vehicle for this mediation to happen. Since our scenario was in an urban environment, we only included events in which drivers had a THW lower than 6 s. This was based on the method used by Vogel (2002), who found that a 6 s THW was the optimal threshold for distinguishing between free, and following vehicles, in urban environments.

9.3.2 Statistical analyses

We used Kolmogorov-Smirnov tests to assess the normality of the data. Whenever the normality assumption was violated, we used logarithmic transformations to correct the observed positive skew, allowing the use of parametrical tests. If transformations were applied, the results of the statistical tests shown are based on the transformed data, but the plots and graphs are generated using the untransformed data.

We analysed data with SPSS V.24 (IBM, Armonk, New York, USA), and generated the visualisations in R. An α-value of 0.05 was used as the criterion for statistical significance, and partial eta-squared was computed as an effect size statistic. Unless otherwise stated, variance of the data was homogenous, as assessed by Levene's test of equality of error variance. Similarly, following log transformation of the skewed data, covariance of the data was homogenous, as assessed by Box's test of equality of covariance matrices.

9.4 Results

9.4.1 Mean time headway

To understand the characteristics of the underlying car-following behaviour, we first plotted the THW distributions during car-following for each condition. Figure 9.3 shows all car-following events for the L2 and L3 groups, for the Baseline Manual Drive, and the manual drives after the Long and Short THW conditions, and Car and No Car conditions. The 6 s threshold we employed seemed to separate car-following from free driving scenarios, as there were no outliers across the distributions. The THW distributions generally followed the distributions observed in other studies, except for the Baseline Manual Drives and the post-automation drives in which drivers experienced a Long THW and resumed control when there was no lead vehicle. Here, longer THWs were generally observed compared to the other drives.

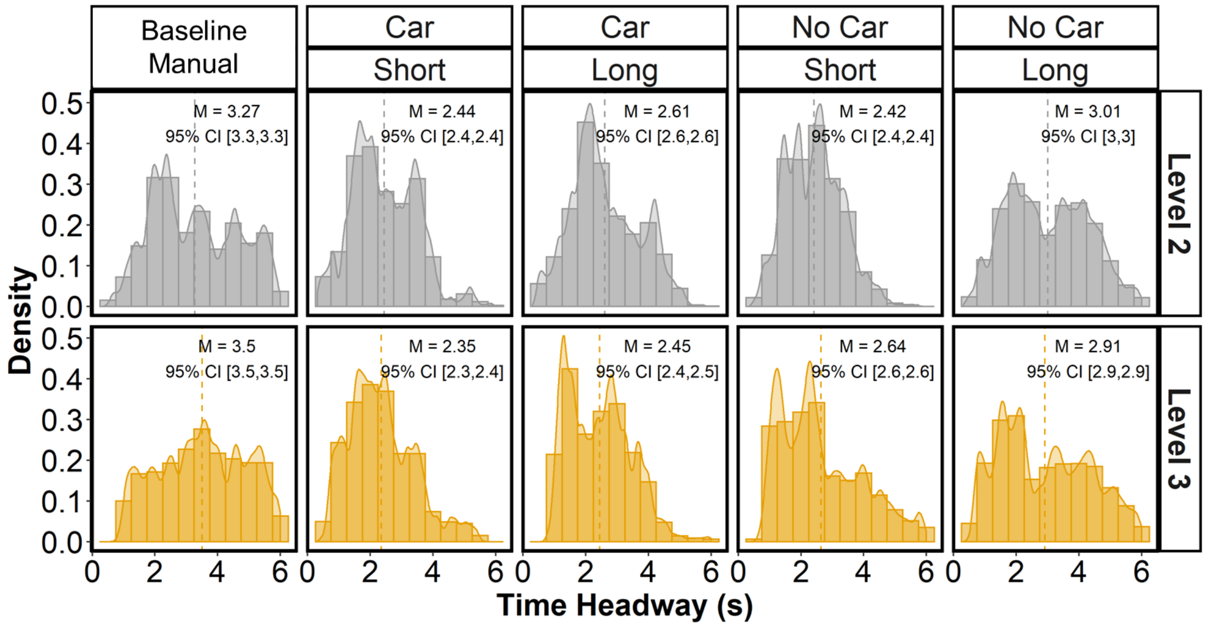

Figure 9.3: Time headway distribution for the L2 and L3 groups, for the Baseline Manual Drive and each of the Long and Short conditions, and Car and No Car conditions. Vertical dashed lines represent the distribution mid-points.

Since there was only one Baseline Manual Drive per participant, and four post-automation manual drives, it was not possible to assess, in a single step, whether there were changes in THW, after each automation drive. Therefore, the analysis of mean THW changes was conducted in two parts: First, we compared drivers' THW in the Baseline Manual Drive with each combination of Time headway and Take-over type conditions, using four separate 2X2 ANOVAs. For each analysis, we used a within-participant factor of Exposure to automation (Baseline Manual Drive, Post-Automation Drive), and a between-participant factor of Level of Automation (L2, L3). The Post-Automation Drive was based on the specific combination of conditions drivers were exposed to. For example, if they resumed control in the presence of a lead vehicle after a Short THW condition, this is referred to as “Car+Short”. The same applies to the other condition combinations: “No Car+Short”, “Car+Long”, “No Car+Long”.

Second, to understand whether changes were influenced by any of the experimental conditions, we calculated the difference in THW between Baseline Manual Drive and the post-automation manual drives, and then compared these using a 2X2X2 mixed ANOVA. The within-participant factors were Time headway during automation (Short, Long) and Presence of lead vehicle during Take-over (Car, No Car), and the between-participant factor was Automation condition (L2, L3). This was used to investigate whether drivers changed their THW after being exposed to automation, and if the conditions influenced the magnitude of this change. Initially, we included each of the subscales of AISS, T-LOC, and DSQ as covariates in the ANOVAs. However, all of these sub-scales returned non-significant effects and small effect sizes. Therefore, to maintain statistical power, these covariates were removed from the analyses.

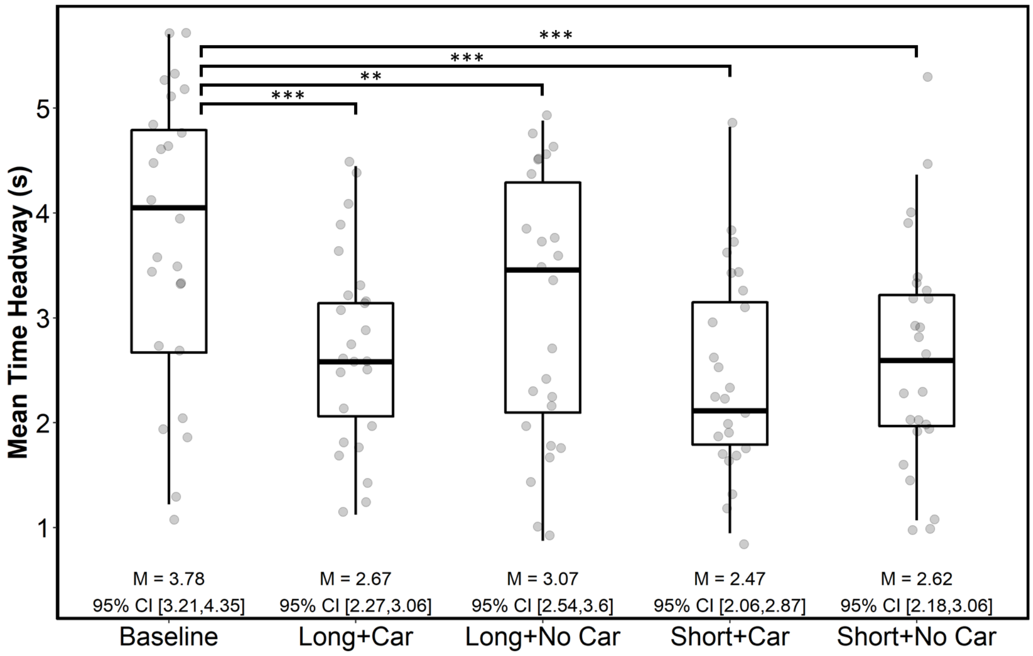

*Figure 9.4: Mean time headway (s) during manual car-following during the Manual Baseline Drive and the four post-automation manual drives. ** $p < .005$ *** $p < .001$*

The first set of ANOVAs we conducted revealed that drivers’ THW in the Baseline Manual Drive was significantly higher, compared to all subsequent post-automation manual drives (Figure 9.4). On average, in the Baseline Manual Drive, drivers had a THW of 3.78 s, whereas the global mean for all post-automation car-following events was 2.7 s. In other words, absolute THW during car-following decreased significantly after experiencing automated car-following. Across all ANOVAs, there was no effect of Level of Automation and no interactions, which suggests that the reduction in THW occurred irrespective of whether drivers were engaged in an NDRA during automation (L3), or were looking around the road environment during automation (L2).

The second ANOVA revealed that there was a main effect of length of Time headway during automation ($F(1,23)=4.320$, $p<.05$, η_p^2 =.158) on how much drivers changed their THW, compared to their Baseline Manual Drive. As shown in Figure 9.5A, drivers had significantly shorter THWs during the post automation manual car-following, after the Short THW conditions (M: -1.25 s), compared to after the Long THW conditions (M: -.9 s). Therefore, there was an immediate effect of the set THW during automation, on drivers' subsequent adopted headway. There was also a main effect of presence of lead vehicle during take-over ($F(1,23)=11.339$, $p<.01$, η_p^2 =.330), where Figure 9.5B shows THW during post-automation car-following appeared to reduce significantly more, relative to the Baseline Manual Drive, if drivers resumed control in the presence of a lead vehicle (M: -1.23 s) compared to resuming control without a lead vehicle (M: -.92 s). In other words, we know that drivers reduce their THW during car-following after experiencing automated car-following, but the reduction is more pronounced if drivers resume control during a car-following event, rather than restarting a car-following event a little later. These results suggested that drivers were not only mimicking the THW they had just experienced, but the effect was more pronounced when the car-following event persisted through the resumption of control.

There was no effect of Automation level ($F(1,23)=.006$, $p=.999$, η_p^2=.000) and no interactions, which suggests that monitoring the environment and observing the THW during automated car-following (L2) did not influence the extent to which drivers reduced their THW. While the NDRA in the L3 group was designed to take drivers' visual attention away from the forward path, it is possible that they made short glances to the road during automation. In this case, the results suggests that ADF use can influence drivers' behaviour, even if they are not fully aware of, or continuously monitoring, its performance. In addition, all drivers were exposed to the lead vehicle for a short period immediately after the TOR, which may also have influenced their subsequent adopted headway.

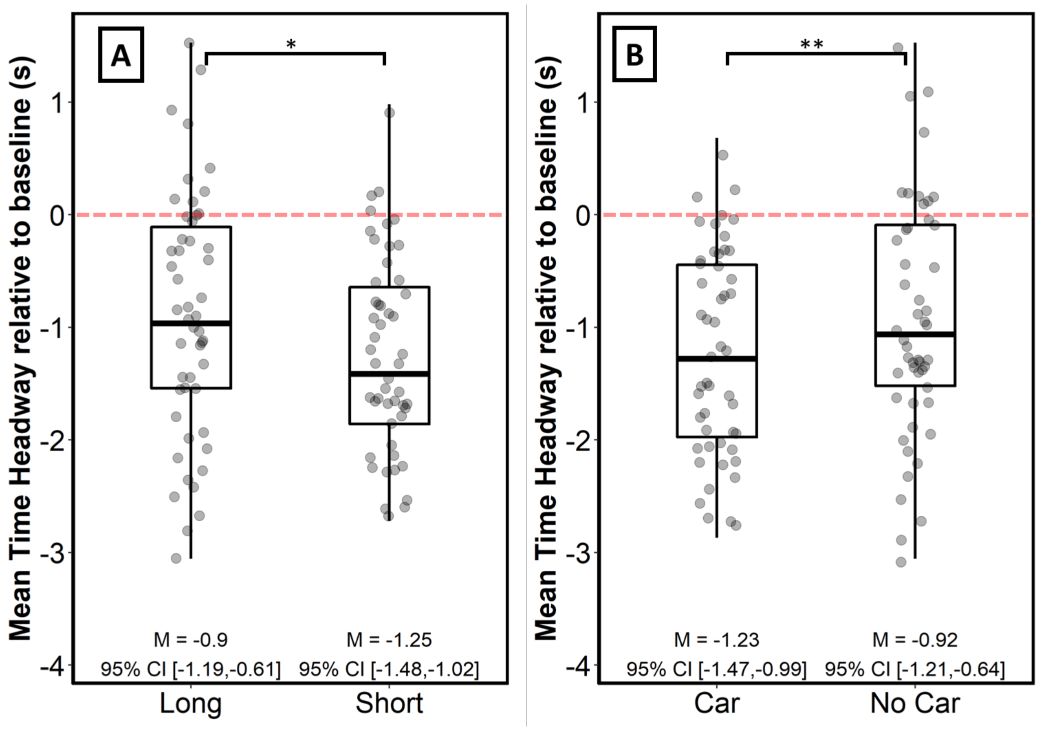

*Figure 9.5: Difference in mean time headway (s) during car-following between the Baseline Manual Drive and post-automation drives, for the conditions where A) drivers experienced Long or Short THW during automated car-following, and where B) drivers resumed control with a lead car (Car) or without a lead car (No Car). The red dotted line represents the Baseline Manual Drive for all drivers. * p < .05 ** p < .01*

9.4.2 Standard deviation of time headway

One of the primary concerns about the effect of vehicle automation on drivers' behaviour is the extent to which it affects their control of the vehicle, once they resume manual control. Mean THW is a useful measure for understanding the degree of risk that drivers are willing to accept during car-following. However, equally important, from a controllability standpoint, is the steadiness or consistency with which drivers control their vehicle after automation. During car-following, this would be reflected by the variation in drivers' THW, which also indicates drivers' 'safety boundary' (Boer, 1999). To examine whether there were changes to the variation in drivers' THW, we followed the same two stages of analysis described above. First, we compared drivers' standard deviation (SD) of THW in the Baseline Manual Drive with each combination of Time headway and Take-over type, using four separate 2X2 ANOVAs, with a within-participant factor of Exposure to automation (Baseline Manual Drive, Post-Automation Drive - as described in the previous section), and a between-participant factor of Level of Automation (L2, L3). Second, to understand whether changes were influenced by the Time headway and Take-over type, we calculated the difference in SD of THW between Baseline Manual Drive and the respective conditions, comparing these with a 2X2X2 ANOVA. The within-participant factors were Time headway during automation (Short, Long) and Presence of lead vehicle during Take-over (Car, No Car), and the between-participant factor was Automation condition (L2, L3).

Across all four 2X2 ANOVAs, comparing SD of THW in the Baseline Manual Drive to post-automation manual drives, there was no effect of Exposure to automation, no effect of Level of Automation and no interactions (Figure 9.6). These results indicate that, while drivers may have reduced their THW in car-following after automated car-following, their behaviour was quite consistent across the different conditions.

The second ANOVA revealed no effects of Exposure to Automation, Presence of lead vehicle during Take-over, Level of Automation, and no interactions, which is not surprising given that absolute SD of THW of each condition did not differ significantly, compared to Baseline Manual Drive. This indicates that THW variability was not influenced by whether drivers had their eyes away from the forward roadway during automation.

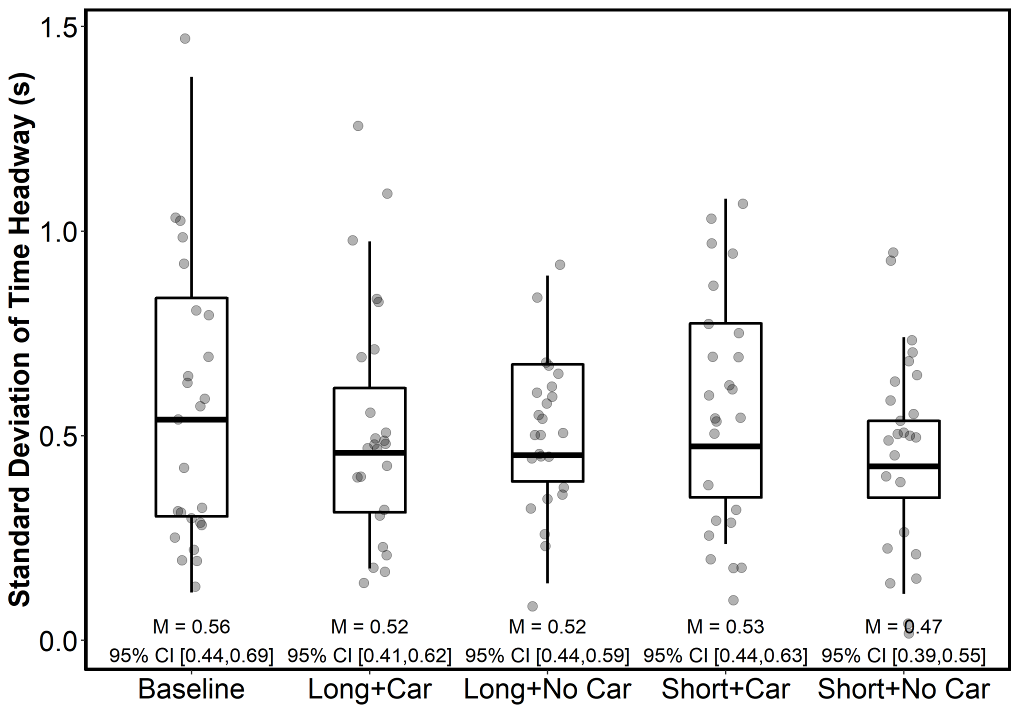

Figure 9.6: Standard deviation of time headway (s) during manual car-following for the Baseline Manual Drive and the four post-automation manual drives.

9.4.3 Subjective assessment

In addition to the T-LOC, AISS, and DSQ questionnaires, drivers were asked to provide a subjective assessment of the ADFs' behaviour, after the automated car-following drives with Long and Short THW (top three questions in Figure 9.7). Drivers were also asked to assess changes in their behaviour after each of these drives (bottom three questions in Figure 9.7).

76% of drivers felt that the ADF kept a safe distance from the car in front, during the Long THW condition, while 84% of drivers disagreed with this statement for the Short THW condition. For both the Long and Short THW conditions, most drivers (84% and 92%, respectively) did not feel that the ADF should have kept a closer distance from the car in front. However, there was more consensus across drivers that the ADF should have kept a longer distance to the lead vehicle, for the Short THW condition. These responses suggest that drivers were able to differentiate between the experimental conditions, and while the Long THW condition was generally tolerable, the Short THW was viewed as unsafe.

There was no clear agreement between drivers about whether they had changed their behaviour after using the ADF, though most drivers felt that they did not use the brakes and accelerator pedals more after the automated drives. Given that drivers assessed the Short THW condition to be unsafe, it is unsurprising that 92% indicated that they did not keep the same distance to the lead vehicle in the subsequent manual drive.

To determine whether what drivers' subjective response is in terms of their perceived behaviour after automation was reflected in their actual behaviour, we ran two separate Pearson product-moment correlations of drivers' responses, comparing response to the item "*Experiencing the automated driving system changed how I drove in the subsequent manual drive"* with their actual mean THW. Post automation THW was compared to Baseline Manual Drive values, for both the Long and Short THW conditions, while also controlling for the Level of Automation.

There was a moderate, negative significant correlation between the two measures for the Long THW condition ($r(47)$=-.341, p=.016), showing that what drivers thought they did was opposite to what they actually did. However, there was no significant association in the Short THW condition ($r(47)$=-.010, p=.944). Therefore, drivers' assessment of their behaviour did not match their actual behaviour.

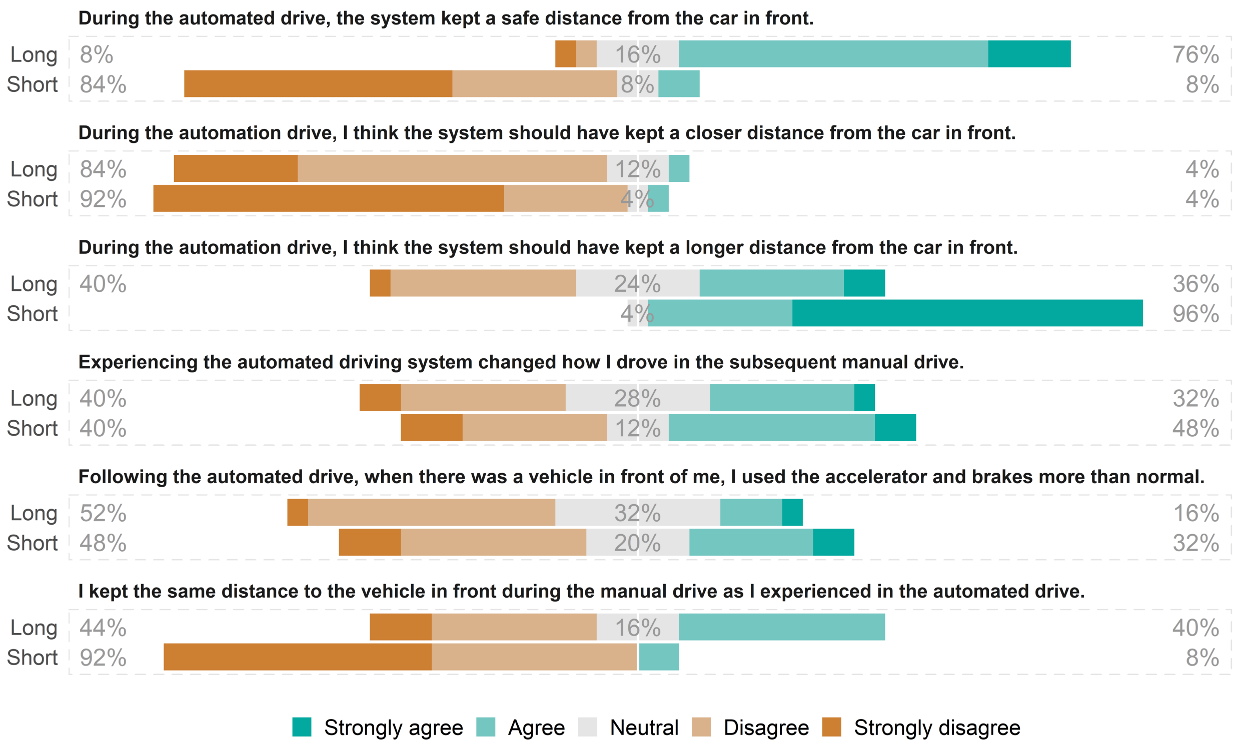

Figure 9.7: Drivers' subjective assessment of the ADFs' behaviour during automated car-following, and their judgement of their own behaviour during post-automation manual driving.

9.5 Discussion and Conclusions

This driving simulator study assessed changes in driver's manual car-following behaviour after automated car-following in an urban environment. The study had two experimental groups: during automated car-following, one group was engaged in an NDRA (L3), while the other group was free to look around the road environment (L2). We also compared the effect of Long (1.5 s) and Short (.5 s) THW conditions during automated car-following, and whether the presence of a lead vehicle, during the resumption of control, had an impact on any subsequent changes in car-following behaviour. All post-automation drives were compared to a Baseline Manual Drive, which was recorded at the start of the experiment.

As our first research question, we sought to understand whether drivers change their car-following behaviour in manual driving after experiencing car-following in automated driving. Our results showed that drivers significantly reduced their time headway in all post-automation drives, compared to a Baseline Manual Drive. This is in line with the findings of both Skottke et al. (2014), Eick & Debus (2005), who showed that drivers reduced their time headway after being decoupled from highly automated driving and truck platoons. This pattern has also been observed in a study on drivers' behavioral adaptation after using full-range ACC (Varotto, 2020). This can be explained through risk homeostasis theory, where, as drivers become more familiar and comfortable with shorter THWs during automated driving, they adjust their boundary of acceptable risk. In other words, drivers become used to following at shorter distances with no negative outcomes, despite not being in control of the vehicle. However, as drivers' resume manual control, this adapted risk boundary carries over into their own manual driving, and they accept shorter THWs than they otherwise would. The observed changes in behaviour justify our concern regarding the potential increased susceptibility to rear-end collisions after automated driving, as shorter THWs increase the risk of rear-end collisions (Lee, Llaneras, Klauer, & Sudweeks, 2007). Future research should confirm our results and examine the extent to which adaptation of car-following behaviour after automated driving impacts drivers' abilities to respond in such situations.

Our second and third research questions addressed whether any changes in post-automation car-following was influenced by the THW adopted by the automated driving system and whether drivers resumed control in the presence of a lead vehicle. Our results showed that there was a greater reduction in THW after drivers resumed control in the presence of a lead vehicle, and also after they had experienced a shorter THW (0.5 s) during automated car-following. These results demonstrate that the THW drivers adopt in manual car-following is influenced by the THW they were exposed to during automated car-following, especially if the car-following event persists through the resumption of control. While shorter THWs adopted by automated vehicles may lead to optimised traffic flow and capacity (Friedrich, 2016), our results suggest that this should be carefully balanced against the potential negative impact this will have on drivers' manual driving behaviour, as well as their acceptance and, ultimately, use of the system.

For our final research question, we sought to understand whether any changes in post-automation car-following would be influenced by whether or not drivers engaged in a visual NDRA during automation. We found that there were no differences in THW changes between the L2 and L3 groups, suggesting that drivers do not need to continuously monitor the road environment for their THW to be influenced by the ADF behaviour. It could be that during L3 driving, drivers perceived the lead vehicle via peripheral vision, possibly reinforced by short glances to the roadway. However, future research should clarify this hypothesis.

Based on research by Itkonen & Lehtonen (2020) and Rudin-Brown & Parker (2004), another aim of this research was to investigate whether any changes in behaviour were associated with drivers' self-reported traits, including sensation seeking (AISS, Arnett, 1994), traffic locus of control (T-LOC, Özkan & Lajunen, 2005), and driver style questionnaire (DSQ, French, West, Elander, & Wilding, 1993). However, the changes in THW we observed did not appear to be associated with any subscales of the T-LOC, AISS, or DSQ questionnaires, suggesting that the changes observed here may not be linked to the underlying personal traits we investigated. These results contrast with previous work on the link between individual characteristics such as sensation seeking and locus of control on behaviour changes (Ward, Fairclough, & Humphreys, 1995; Rudin-Brown & Parker, 2004).

In addition to the above, we sought drivers' perceptions of their own and the ADF behaviour during car-following. Drivers' subjective responses showed that their change in behaviour was not necessarily reflected in their subjective assessment of their behaviour change. That is, drivers were not aware that they had changed their behaviour, even in the short THW condition, which they overwhelmingly rated as unsafe. This is not surprising, as previous studies have shown that individuals are not always aware of how the use of technology can change their behaviour, for example, the effect negative effect of using a mobile phone while driving on performance and mental processing (Boase, Hannigan, & Porter, 1988; Alm & Nilsson, 1995).

9.6 Recommendations

Our research highlights a number of areas that can be addressed to limit the adverse effect of BA to automation on manual driving. First, the system in use should be designed in a way that limits negative BA. For example, it is clear from drivers' behavioural change in the current study, that the system should have adopted a more conservative THW. Second, drivers should receive explicit training about the potential effects that automation use may have on their manual driving, so that they do not become complacent. For example, if in the current study drivers were warned that their THW might shorten after using automation, it may have reduced the likelihood that this occurred. Third, drivers should be warned when their behaviour exceeds certain safe boundaries of operation. For example, in the current study, drivers could have been warned during manual driving that their THW had shortened compared to either their normal driving style or a safe standard.

9.7 Limitations

We should note that the THWs adopted by drivers in the first manual drive of this experiment is longer than what is commonly observed during real-world car-following. This may be due to drivers' unfamiliarity with the driving simulator and the urban road environment they were travelling in. If the THW observed in the Baseline Manual Drive is higher than that adopted by our participants in real-world driving, it may partially account for the reduction in post-automation THW. However, this would not account for the differences observed between the Long and Short THW conditions, or the Car and No Car conditions. Moreover, Vogel (2002) found that 6 s THW is an optimal threshold for distinguishing between free and following vehicles in urban environments, suggesting that the behaviour we observed in this study can be considered to be car-following and not free driving. In addition, the experimental drives used in this study were relatively short, and though it is interesting to note that behavioural adaptations may exist after such a short period, this may not necessarily represent the real-world pattern system usage. For example, we did not consider the impact that fatigue and hypovigilance may have had on drivers' attention to the car-following task during automated driving, and, therefore, on their car-following behaviour in subsequent manual car-following.

9.8 Future work

Notwithstanding the above concerns, the trends observed here are generally in line with those of Skottke et al. (2014) and Eick & Debus (2005), who found that drivers reduced their THW for periods in manual driving, after decoupling from fully automated driving. However, it is an open question whether the kinds of changes we observed here would be seen after the use of ADF in daily use. For example, how does drivers' behaviour change after using ADF over more extended periods, such as weeks or months? It is also important to consider whether behavioural adaptations are consistent across different settings, for example, on motorways, rural roads, and urban environments. Furthermore, are there behavioural adaptations after using ADF in different use-cases, such as lane changes, parking, or merging? There is also merit in investigating whether the type of take-overs (i.e., critical vs non-critical; Erikson & Stanton, 2017) influence the extent to which behavioural adaptation carries over into subsequent manual driving. Finally, how does behaviour adapt after different usage patterns, for example, less frequent, but more extended periods vs more frequent, but shorter periods, as previous research has shown that regular use of cruise control, for example, can lead to a reduction in vigilance and increase in reaction time (Dufour, 2014). Therefore, future research should endeavour to investigate these issues, as ADF use will become more widespread in the coming years, and it is imperative that we understand the prospective risks of using ADAS and ADF.

9.9 Conclusions

Our results build on the research into behavioural adaptation and ADAS use and show that there is the potential for drivers' behaviour to adapt after using automated driving systems, during car-following. In the coming decades, humans will likely be still involved in the driving task to varying degrees, so it is important from a safety perspective to understand what issues there are and for researchers and vehicle manufacturers to develop appropriate countermeasures.

10 Driving Simulator Study: Evaluating an ambient peripheral light display in automated driving

In this chapter a driving simulator study is described that explored the impactof HMI-design, in this case of an ambient display on trust and system usage. The study was conducted by the University of Leeds.

10.1 Aim

Ambient LED displays provide peripheral light-based cues to drivers about a vehicle's current state, along with requests for a driver's attention or action. They have been investigated as potential collision warning tools Danielsson et al. (2007), lane change decision aids (Kunze, Summerskill, Marshall, & Filtness, 2019), a means to help modulate drivers' speed (Meschtscherjakov, Döttlinger, Rödel, & Tscheligi, 2015; van Huysduynen, Terken, Meschtscherjakov, Eggen, & Tscheligi, 2017), and to guide drivers' attention to identify targets (road users/obstacles), and indicate vehicle intention (Schmidt & Rittger, 2017; Trösterer, Wuchse, Döttlinger, Meschtscherjakov, & Tscheligi, 2015). Peripheral ambient light displays have also been used to inform drivers of malfunctioning ADAS (Langlois, 2013), and to facilitate collaborative driving tasks between the driver and the co-driver (Meschtscherjakov et al., 2015).

Recently, light displays have been applied in the context of automated driving. For example, Borojeni, Chuang, Heuten, & Boll (2016) conveyed contextual information through ambient displays to assist drivers during take-over requests and found that this resulted in shorter reaction times and longer times to collision, without increasing driver workload. More commonly, light displays have been to provide information/warnings to drivers about other road users, or the AVs intentions (Dziennus, Kelsch, & Schieben, 2016). The research in both manual and automated driving shows that, in general, ambient lights are rated highly by drivers, and drivers are sensitive to peripheral cues (Kunze et al., 2019). However, few studies have investigated the use of these displays to improve drivers' perceptions of trust and safety during automated driving, and to facilitate transitions between L3 automated driving and manual driving.

Therefore, the current driving simulator study addressed this gap and also three of the main research questions of the L3Pilot project. First, we aimed to evaluated the effectiveness of an ambient peripheral light display (Lightband HMI) in terms of its potential to improve drivers' trust in L3 automation, measured through a questionnaire (RQ-U3) and, second, through level of engagement in a non-driving task during L3 automated driving (RQ-U9). Third, we assessed whether this Lightband HMI could be used to facilitate effective transitions of control between L3 automated driving and manual driving, compared to an Auditory alert (RQ-U10).

10.2 Methods

10.2.1 Participants

Following approval from the University of Leeds Research Ethics Committee (Reference Number: LTTRAN-132), we recruited 41 drivers, via an online social media platform. Participant demographic details are displayed in Table 10.1. Participants received £30 for taking part in the experiment and were free to withdraw at any point.

Table 10.1: Participant demographics information.

Demographics	Gender Mean (SD)	
	Males (N=20)	Females (N=21)
Age (years)	44 (13)	44 (13)
Years with licence	25 (13)	24 (12)
Miles driven annually	10300 (5332)	6642 (3350)

10.2.2 Design and Procedure

10.2.2.1 Equipment

The experiment was conducted in the full motion-based University of Leeds Driving Simulator (UoLDS), which consists of a Jaguar S-type cab, housed in a 4m diameter spherical projection dome with a 300° field-of-view projection system. The simulator also incorporates an 8 degree-of-freedom electrical motion system. This consists of a 500mm stroke-length hexapod motion platform, carrying the 2.5T payload of the dome and vehicle cab combination, and allowing movement in all six orthogonal degrees-of-freedom of the Cartesian inertial frame. Additionally, the platform is mounted on a railed gantry that allows a further 5m of effective travel in surge and sway. A Seeing Machines Driver Monitoring System was used to record the participants' eye movements at 60Hz. Inside the simulator's vehicle cabin, a Liliput 7" VGA touchscreen with 800X480 resolution, was installed near the gear shift, and used for a non-driving related, secondary task, described below.

10.2.2.2 Experimental Design

In this experiment, participants compelted two experimental drives. Each experimental drive lasted ~17 minutes, with five ~2-minute automation segments, interspersed with ~1-minute manual driving segments (Figure 10.1). There were five take-over requests per drive, and 10 in total. The entire experiment lasted approximately 2 hours. In borth experimental drives, during automated driving, participants were instructed to engage in a visual non-driving related “Arrows” task (NDRT; Jamson & Merat, 2004). The Arrows task requires participants to search for, and touch, the upward-facing Arrow, displayed in a 4x4 grid of Arrows, using a touch screen in the centre console (see Figure 10.2). Each time the upwad-facing arrow is correctly idenfied and selected, a new grid of arrows is generated. The screen displayed the current participant’s cumulative score and a ‘score to beat’ to keep them engaged in the task.

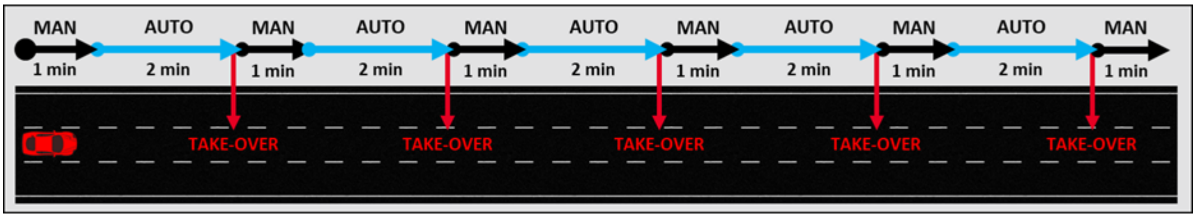

Figure 10.1: Schematic representation of each experimental drive.

A 2X5 within-participant design was used for this study, with the factors *HMI type* (Lightband, Auditory) and *Take-over number* (1-5). HMI type was fully counterbalanced across participants.

HMI type specifies the HMI drivers were presented with during automated driving, and used for the take-over i.e. Lightband or Auditory. In terms of the instructions provided to drivers for take over, the same text and symbols were displayed in the vehicle's dashboard display (HMI) for both conditions (Figure 10.4).

In the *Lightband condition*, an LED-based lightband notification system was displayed in the vehicle cabin during automated driving and for signaling take-overs (Figure 10.3). During manual driving, the lightband was not active. When automation was available to be engaged, the lightband pulsed with a blue light at 2 Hz until the driver turned automation on. During automated driving, the lightband displayed a solid blue light to indicate that the automation was operating normally. During take-over requests, the lightband pulsed with a red light at 2 Hz until the driver resumed manual control. The Lightband HMI was not accompanied by any auditory warnings.

In the *Auditory condition*, participants received an auditory alert (880 Hz, lasting 0.2 s) to notify the driver to engage or disengage the automated driving system.

Take-over number specifies the number of times drivers resumed control during the experimental drive, for each HMI condition.

Figure 10.2: Example of a driver performing the Arrows task during automated driving in the Lightband HMI condition.

Figure 10.3: Example of the Lightband HMI during automated driving (left), and the placement of the automation status symbol and vehicle speed in the dashboard display (right). In both figures, the Driver Monitoring System is located above the on the dashboard above the steering sheel.

10.2.2.3 Automated Driving System and Human-Machine Interface

When active, the automated driving system (ADS) assumed lateral and longitudinal vehicle control and maintained a maximum velocity of 70 mph. The status of the ADS was indicated through a symbol that was located on the left panel of the vehicle's dashboard display (Figure 10.4; Human-Machine Interface). The symbols for "Take-over request" and "Engage automation" pulsed at a rate of 2 Hz until the driver resumed control or engaged automation as required. The display of the symbols for "Manual control" and "Automation engaged" remained constant.

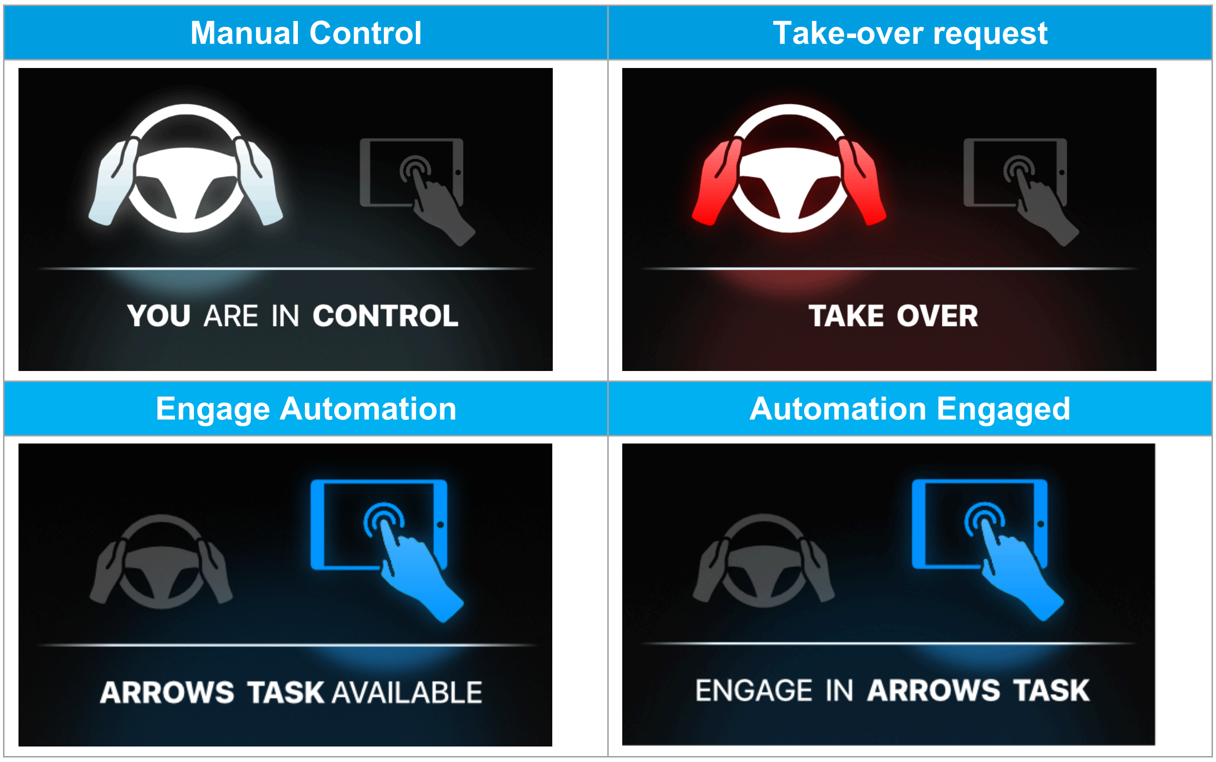

Figure 10.4: An example of the HMIs located in the vehicle's dashboard display.

10.2.2.4 Procedure

During recruitment, participants were emailed a screening and demographics questionnaire, which included questions about age, gender, driving experience, experience with different Advanced Driver Assistance Systems (ADAS). The questionnaire also included the traffic locus of control questionnaire (T-LOC; Özkan & Lajunen, 2005). To be eligible to take part in the experiment, participants had to hold a valid licence to drive a car, have at least one year's experience driving in the UK, and not have participated in a driving simulator study that included interaction with automated vehicles. Prior to arrival, participants were emailed a description of the study, information about COVID-19 procedures during the experiment, and were asked to sign a consent form.

Upon arrival at the simulator, the experimenter asked the participant a series of questions to ensure COVID-19 compliance. They were then taken into the building where the experiment was explained in more detail, and they were given the opportunity to ask questions

Participants were taken into the simulator dome and the experimenter explained all the safety procedures, driving controls of the vehicle, and various dashboard icons, as well as how to do the Arrows task, as well as engaging and disengaging the automated driving system. The drives took place on a three-lane motoway with ambient traffic. To enable automation, participants were asked to drive in the centre of the middle lane and maintain the 70-mph speed limit and adhere to the standard rules of the road, ensuring safe operation of the vehicle, throughout the drive. Before each of the two experimental drives, participants performed a short practice drive. To avoid confusing participants by showing them both HMIs at the start of the experiement, they were only shown the HMI system that they would experience in the subsequent experimental drive (i.e., Lightband or Auditory HMI).

They were then left in in the simulator dome to perform the first practice drive to allow them to become familiar with the simulator controls and motion system. The experimenter talked through the practice drive with the participant via an intercom system. After the practice drive, participants remained in the dome and were asked if they were happy to continue to the main experimental drive.

The experiment began with the participant driving in manual mode for a couple of minutes, after which they received an instruction from the automated driving system to turn the automation on. This was achieved by pulling the left indicator stalk towards them. Once automation was engaged, participants began performing the Arrows task. After approximately 2 minutes, participants received a notification to take over control. To turn automation off, participants had to have both hands on the steering wheel (as judged by the touch-sensitive steering wheel), be looking on the road ahead (as judged by the driver monitoring system) and pull the left indicator stalk towards them. There was no lead vehicle or obstacle during the take-overs, however, during the automated drive, other vehicles in the adjacent lane did move in and out of the driver’s lane. Our aim here as to implement a non-critical take-over request.

After the practice drive, and after each experimental drive, participants rated their perceptions of trust, safety, and HMIs, by answering a series of questions listed in Table 10.2.

Table 10.2: Post-experiment questionnaire.

Drive	Question	Response format
Post-practice	I trust that the vehicle will drive safely, while I do the Arrows task	5-point scale (Strongly disagree-Strongly agree)
Post Drive 1	I trusted that the vehicle would drive safely while I did the Arrows task	5-point scale (Strongly disagree-Strongly agree)
	If your level of trust in the automated driving system changed since the start of the experiment, please explain why.	Free text
	I felt safe while doing the Arrows task during automated driving	5-point scale (Strongly disagree-Strongly agree)
	In this drive, the Ligtband/Auditory signal was...	5-point scale for each Van der Laan Scale item
Post Drive 2	I trusted that the vehicle would drive safely while I did the Arrows task	5-point scale (Strongly disagree-Strongly agree)
	If your level of trust in the automated driving system changed since the start of the experiment, please explain why.	Free text
	I felt safe while doing the Arrows task during automated driving	5-point scale (Strongly disagree-Strongly agree)
	In this drive, the Ligtband/Auditory signal was...	5-point scale for each Van der Laan Scale item
	How engaged were you with the arrows task while automation was on?	10-point (Not at all engaged-Highly engaged)
	Apart from take-over requests, was there anything that interrupted your engagement in the arrows task while automation was on? If so, please explain briefly.	Free text
	Which warning system did you prefer?	Lightband/Auditory

10.2.3 Statistical analyses

We analysed data with SPSS V.24 (IBM, Armonk, New York, USA), and generated the visualisations in R and Microsoft Excel. An α-value of 0.05 was used as the criterion for statistical significance, and partial eta-squared was computed as an effect size statistic. Unless otherwise stated, variance of the data was homogenous, as assessed by Levene's test of equality of error variance.

A number of different parametric and non-parametric statistical tests were conducted as part of the analysis. First, we conducted two repeated measures ANOVAs [2 (HMIs) x 5 (Take-over Number)] to investigate the effect of HMIs and Take-over Number on drivers' hands on wheel time and automation disengagement time, respectively. Hands on wheel time was taken from the onset of the take-over request and the point at which both of the drivers' hands were detected on the steering wheel. Automation disengagement time was taken from the onset of the take-over request and the point at which the driver disengaged the automation and the driving mode turned to manual.

In addition to objective measures and drivers' take over performance, drivers' subjective experience was also explored. We conducted a chi-square test to investigate whether drivers' HMI preference was related to the order in which it was experienced. I would change to "A 1X3 repeated-measures ANOVA was conducted on mean perceived trust scores (rated before the first drive, after the Auditory HMI, and after the Lightband HMIs).

A Kruskal Wallis test was used to compare drivers' ratings of perceived safety after experiencing the Auditory HMI and Lightband HMI. Finally, two paired-sample t-tests were used to compare the results from the van der Laan's Usefulness and Satisfying Scale.

10.3 Results

10.3.1 Reaction time

For hands on wheel time, there was no effect of HMI type ($F (1,29) = 3.443$, $p = .074$, $\eta p^2 = .106$), though the mean was higher for the Lightband condition ($M = 2.16$ s, $SD = .14$), compared to the Auditory condition ($M=1.93$ s, $SD= .07$). There was no effect of Take-over number ($F (4,29) = 1.553$, $p = .19$, $\eta p^2 = .051$), and no interactions ($F (4,116) = 1.265$, $p = .228$, $\eta p^2 = .042$).

Similarly, for automation disengagement time, there was no effect of HMI type ($F (1,38) = .364$, $p = .55$, $\eta p^2 = .010$), where the means for the Lightband ($M = 3.46$ s, $SD = .15$) and Auditory ($M = 3.36$ s, $SD = .14$) conditions were similar. There was no effect of Take-over number ($F (4,38) = .284$, $p = .88$, $\eta p^2 = .007$), and no interactions ($F (4,152) = .302$, $p = .867$, $\eta p^2 = .008$). It is important to note that the delay seen in drivers' automation disengagment, which was after they placed their hands on the wheel, reflects the time taken to meet the algorithm used by the driver monitoring system: i.e. that both hands were on the wheel and eyes were looking ahead. Video analysis confirmed that in the three cases where automation disengagement was longer than 10 s, those particular individuals struggled to meet these requirements.

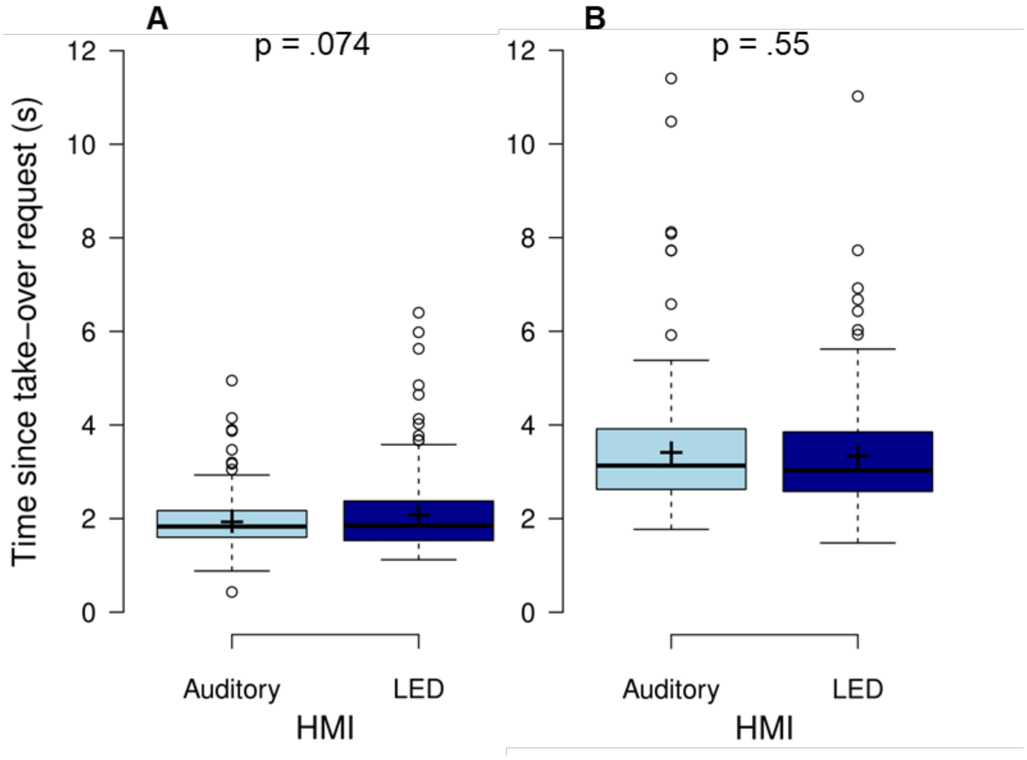

Figure 10.5: Time taken from when take-over request was issued to A) both hands on wheel, and B) automation disengagement, for each HMI. Centre lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots.

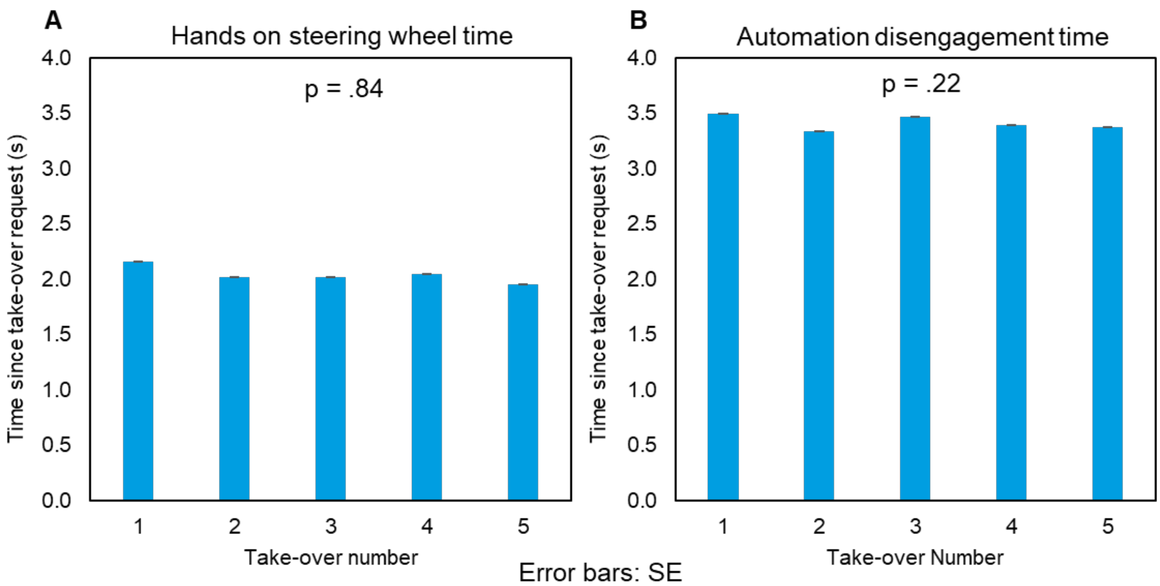

Figure 10.6: Time taken from take-over request was issued to A) both hands on wheel, and B) automation disengagement, for each take-over. Error bars represent Standard Error.

10.3.2 HMI preference

When asked which HMI participants preferred after having experienced both, slightly more participants indicated a preference for the Auditory HMI (54%, N = 22) vs the Lightband HMI (46%, N = 19). However, their preference seems to be have been dependent on which HMI they experienced first (Figure 10.7), $X^2(1, 41) = 5.71$, $p = 0.017$. The most cited reason for participants' preference was that the HMI was more noticeable. However, this reason was given in relation to both the Auditory HMI (N = 12), and the Lightband (N = 7).

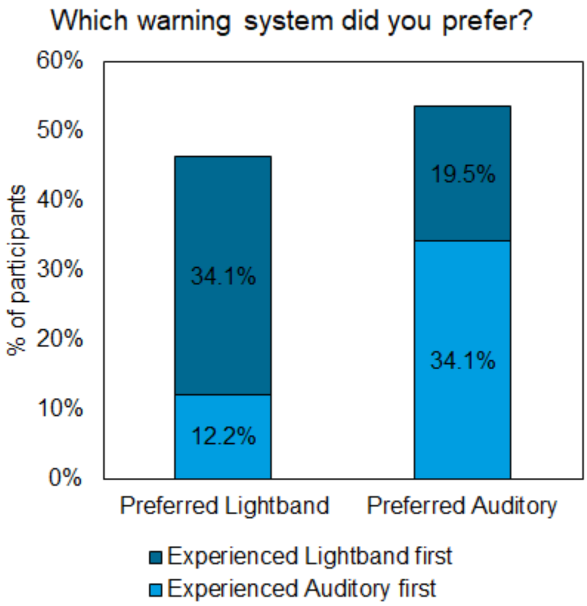

Figure 10.7: Participants' HMI preference.

10.3.3 Perceived Trust

Participants were asked to provide a perceived trust rating on a 5-point scale (Strongly disagree to Strongly agree). before the experiment, and after each drive. The statement they were asked to rate was 'I trust/trusted that the vehicle will drive safely, while I do/did the Arrows task'. A repeated measures one-way ANOVA was conducted to compare the three perceived trust ratings between the baseline, after experiencing the Auditory HMI, and after experiencing the Lightband HMI. Findings showed no significant main effect ($F(2,78) = 2.86$, $p = .084$, $\eta p^2 = .068$; Figure 10.8).

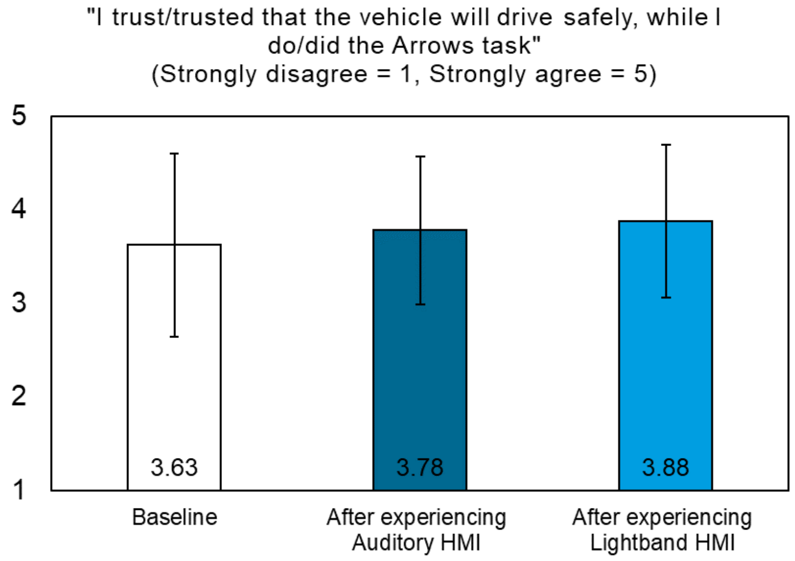

Figure 10.8: Mean scores of participants' responses to the statement "I trust/trusted that the vehicle will drive safely, while I do/did the Arrows task".

10.3.4 Perceived Safety

Participants were asked to rate 'I felt safe while doing the Arrows task during automated driving' on a 5-point scale (Strongly disagree to Strongly agree, Figure 10.9). Overall, participants felt safe during automated driving, but there was no difference between HMIs (Z (1, 40) = 0.37, p = .713).

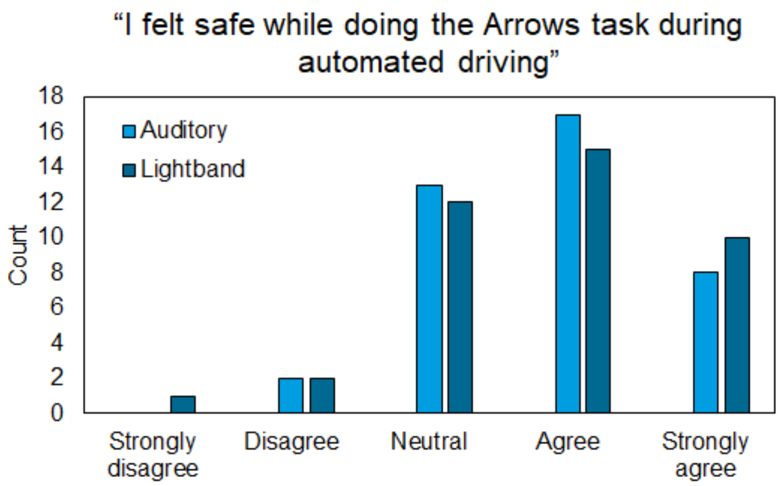

Figure 10.9: Distribution of participants' responses to the statement "I felt safe while doing the Arrows task during automated driving".

10.3.5 Van der Laan scale

Participants' responses to the Van der Laan Usefulness scale showed that both systems were rated as generally useful and satisfying. We compared the responses for each HMI, grouped according to the two constructs within the scale: Usefulness (Useful/Useless, Good/Bad, Effective/Superfluous, Assisting/Worthless, Raising alertness/Sleep-inducing) and Satisfying (Pleasant/Unpleasant, Nice/Annoying, Irritating/Likeable, Undesirable/Desirable). The Auditory HMI (M = 3.37, SD = 0.61) was rated as significantly more useful than the Lightband HMI (M = 3.11, SD = 0.89; t (40) = 2.13, p = .039), but there was no difference in the satisfying scale between the Auditory HMI (M = 3.87, SD = 0.86) and the Lightband HMI (M = 3.66, SD = 1.06; t (40) = 1.52, p = .136; Figure 10.10).

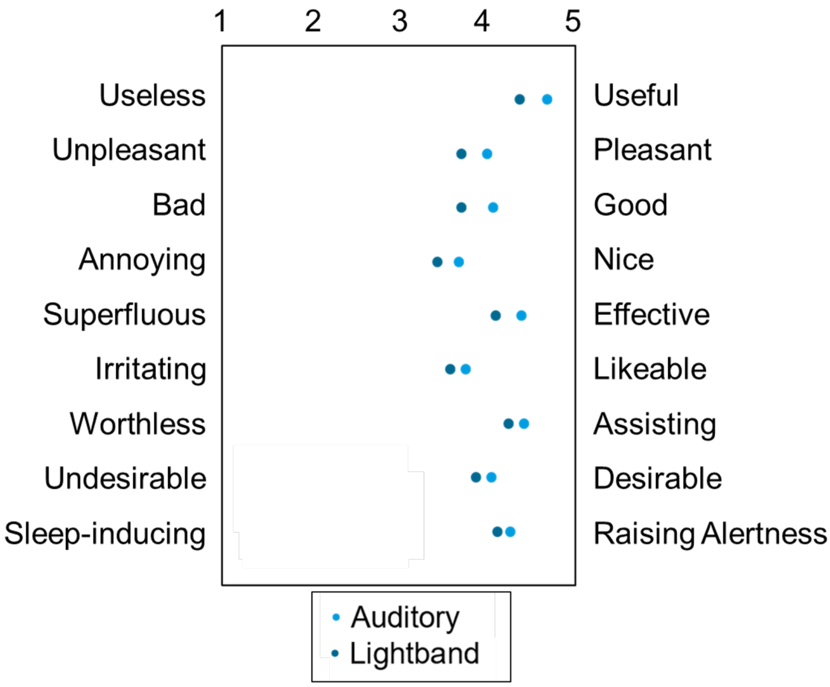

Figure 10.10: Mean scores of participants' responses to the items on the Van der Laan scale, for the Lightband and Auditory HMIs.

10.4 Conclusions

Experience with either the Lightband or Auditory HMI did not appear to improve drivers' perception of safety or trust in the automated driving system during automation, compared to their baseline ratings. There also appeared to be no differences between the HMIs in terms of ratings of safety and trust. Ratings suggest that participants were positively disposed to trust the automated system prior to participation, and the variance in HMIs did not have any impact on this.

Slightly more participants expressed a preference for the Auditory HMI over the Lightband HMI. However, this was significantly impacted by which HMI they had experienced first, suggesting that once drivers experience one form of communication they consider to be effective, they may be slightly resistant to changing it. During the take-overs, participants had marginally faster hands-on wheel time when using the Auditory HMI, compared to the Lightband HMI. In addition, the van der Laan scale results showed that participants rated the Auditory HMI as significantly more useful compared to the Lightband HMI, though no more satisfying. Taken together these results suggest that the Auditory HMI may be slightly more effective in terms of encouraging people to re-take control from L3 automation. However more research incorporating different types of take-over requests is required to gain a further understanding of this issue.

11 Online Study on User Acceptance and NDRA Engagement

In this chapter an online-survey is described that explored drivers' acceptance towards automated driving and their willingness to engage in various NDRAs while driving with an L3-ADF. The study was conducted by the Federal Highway Research Institute (BASt).

11.1 Background, aim, and research questions

Due to the Corona pandemic, BASt's Wizard of Oz study on long-term user acceptance and trust in automated vehicles had to be stopped. BASt reorganised its study approach in order to match test conditions with the restrictions imposed by the pandemic.

Data from tests which could be collected before the pandemic shows an often-varying behaviour of the participants in the car. Especially participants' involvement in non-driving related tasks (NDRA) while driving in automated mode poses questions for further investigation: Level-3-vehicles require the driver to take over control in a timely manner after a take-over request (TOR). Requests may come frequently and unexpectedly and urge the driver to stop performing a NDRA. Since engaging in a NDRA is one of the major promises of Level-3-vehicles, frequent interruptions could possibly lower the acceptance of those vehicles. Therefore, BASt conducted an online study which focused on how participants' acceptance of automated driving may be affected by interruptions of NDRAs due to take-over requests, and which NDRAs are accepted or not accepted with regard to interruption.

The Research questions therefore were:

- RQ-U3: What is the user acceptance of the ADF?
- RQ-U9: What is drivers' secondary task engagement during ADF use?

11.2 Study procedure and methods

11.2.1 Recruitment of participants

Participants for the online study were recruited on the Instagram account of BASt and by sending a recruitment e-mail to BASt employees and other interested persons. They had to write an e-mail to apply and received a personalised invitation link to the study. Personalised links were chosen in order avoid multiple participations by the same person.

11.2.2 Study procedure and questionnaires

After giving their informed consent on data protection and privacy, the participants completed a questionnaire on demographics, their experience as a car driver, their professional background, and a self-rating of their pre-existing knowledge on automated driving. After that, a short text explained SAE Level 3 automated driving including the abilities of the system, the duties of the driver (especially when a TOR is issued: finishing NDRA, reorientation in traffic, confirmation of take-over) and the possibility to perform NDRAs. Being now informed about L3 automated driving, the participants were asked to complete an acceptance questionnaire (van der Laan et al., 1997) with regard to L3 automated driving. Another short text provided instructions on how to complete the following pages of the online study, e. g. to watch each video from start to finish and answer each question from the own perspective.

The main part of the study consisted of nine sites, each with one of nine videos and a set of questions underneath (see below). The sites were presented in a random order in a within-subjects design.

The videos showed a person engaging in an NDRA during an automated car ride and being interrupted by a TOR. The structure was the same for all videos: In the beginning, the NDRA of the video was named on a 5 second still. Then, 15-25 seconds of video followed showing person performing the respective NDRA in a L3 vehicle on the motorway. The person was interrupted by a TOR (warning sound was played aloud and HMI icon was displayed enlarged in the video), then the video froze for ten seconds with countdown depicting the time budget for a safe take-over. Then, the video ended. The shown NDRAs in the nine videos were texting, eating, drinking, making a phone call, app usage, watching a movie on tablet pc, office work on a laptop, gaming on a smartphone and watching the surrounding environment.

Figure 11.1: Still frames from "Texting"-video; left: NDRA engagement, right: TOR with countdown

The set of questions was presented underneath the videos and identical for all videos:

- "Imagine your vehicle was equipped with this function. How often would you engage in [NDRA] while the system is active? (TJM.34)"
  - very often – often – sometimes – rarely – very rarely – never
- „Being interrupted while engaging in [NDRA] would be…"
  - pleasant – unpleasant, nice – annoying, irritating – likeable, undesirable – desirable [satisfaction scale from van der Laan, Heino, & de Waard (1997), used as an indicator for perceived disturbance of the interruption. The usefulness scale from the same questionnaire was not used due to unfit items]

- comfortable – uncomfortable, important – unimportant, controllable – incontrollable, safe – dangerous [items (from Arndt, 2011) for gaining deeper insights into users' perception of interruptions of NDRAs and the following take-over process]

After all nine video sites and associated questions, the van der Laan acceptance questionnaire was presented again to investigate possible changes in acceptance towards L3-automation after having watched the videos.

Finishing the whole questionnaire took approx. 20 minutes. Participants received a compensation of EUR 5.00, BASt employees were excluded from a compensation.

For the analysis of the data, IBM SPSS 25 and Microsoft Excel were used.

11.3 Results

11.3.1 Demographics

A total of 154 participants took part in the study, seven of which had to be excluded due to incomplete data.

The remaining *N* = 147 participants (38.1% female, 61.2% male, 0.7% diverse) were between 18 and 84 years old (*M* = 43.9 years, *SD* = 14.91). They had a university degree (67.3%), vocational training (23.8%) or none of these two (8.8%).

All participants had a driver's licence that they held for less than 1 and up to 65 years (*M* = 25.86 years, *SD* = 14.96) and drove between 1,000 and 80,000 km annually (*M* = 17518.37 km, *SD* = 14096.57).

5.4% of the participants worked for a car manufacturer or supplier, 9.5% developed or researched ADFs, 1.4% tested ADFs and 0.7% are a trained test drivers. Multiple answers were possible. The remaining 85.7% of the participants were not involved in ADF development or the car industry professionally.

11.3.2 User acceptance towards Level-3-automation (RQ-U3)

After being introduced to SAE Level 3 automation, the participants answered van der Laan's acceptance questionnaire in order to assess their acceptance towards Level 3 automation. This was repeated after the presentation of the nine videos to investigate possible changes.

Please note: In van der Laan's questionnaire, negative numeric values stand for high acceptance and vice versa. For better legibilty of the results, all satisfaction and usefulness values were recoded so that positive numeric values depict high acceptance and vice versa.

The following table reports the average scores on the usefulness and satisfaction scale of all participants, male and female participants and four different age groups. The only person of diverse gender is not reported individually due to the small sample size. The age groups were formed with the aim to depict comparable age ranges.

Table 11.1: Uselfulness and acceptance rating before and after NDRA presentation.

.	Usefulness		Satisfaction	
	before	after	before	after
All (N = 147)	.894	.872	1.039	1.027
female (n = 56)	.696	.654	.790	.799
male (n = 90)	1.013	1.011	1.200	1.172
age group 1 (n = 54), ages 18-35	1.011	.941	1.171	1.139
age group 2 (n = 39), ages 36-50	.821	.821	1.045	.936
age group 3 (n = 44), ages 51-65	.791	.823	.881	.978
age group 4 (n = 10), ages 66-84	1.000	.920	1.000	1.000

Both usefulness and satisfaction received fairly high ratings with variations by age and gender groups. These were analysed subsequently:

Paired t-tests ($p < .05$) were conducted in order to investigate changes in usefulness or satisfaction ratings (before/after-comparison) within the mentioned groups: no statistically significant changes were found.

Male participants rated automated driving significantly as more useful and more satisfying than female participants in both, before and after assessment (t-test results: usefulness before: $t(144) = 2.581$, $p = .011$, $\eta_p^2 = .044$; usefulness after: $t(144) = 2.809$, $p = .006$, $\eta_p^2 = .052$; satisfaction before: $t(144) = 3.037$, $p = .003$, $\eta_p^2 = .06$; satisfaction after: $t(144) = 2.692$, $p = .008$, $\eta_p^2 = .048$).

A repeated measures ANOVA was performed to investigate the influence of the age group on the rating of usefulness and satisfaction ratings of automated driving: no statistically significant differences were found.

11.3.3 Perception and use of non-driving related tasks (RQ-U9)

For each of the nine presented NDRAs, the participants answered how often they would engage in these if they had a Level-3-ADF in their car. Figure 11.2 shows the percentage distributions.

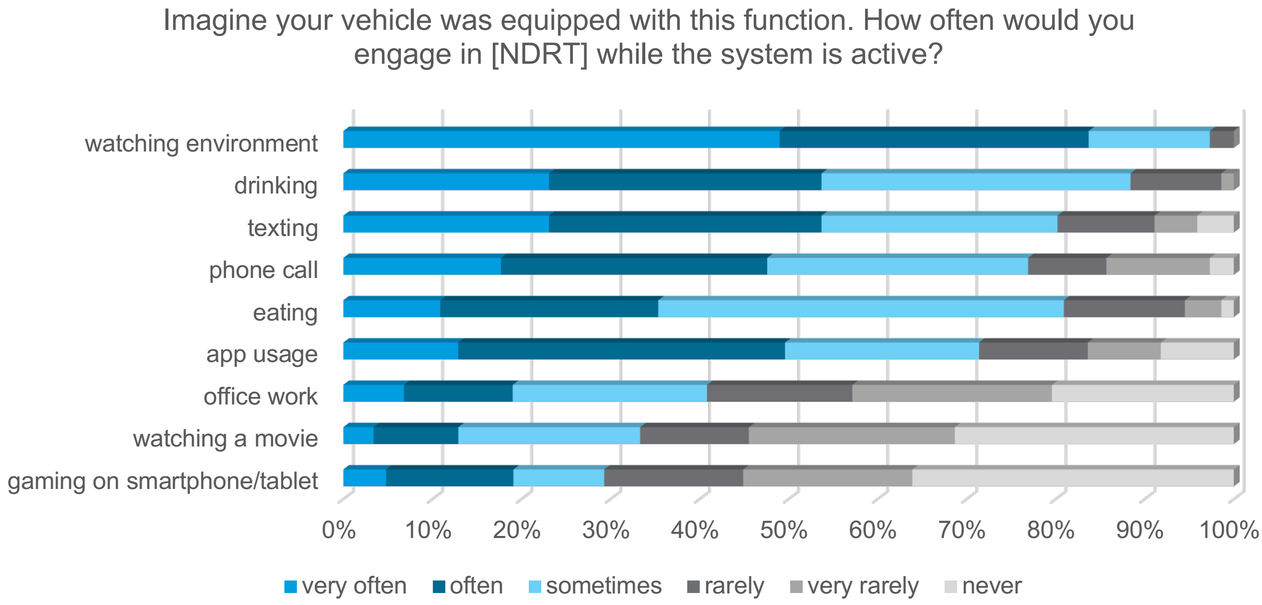

Figure 11.2: Preferred NDRAs, ranked from most-popular (top) to least-popular (bottom).

Over 70% of the participants stated that they would at least “sometimes” engage in *app usage, eating, phone calls, texting, drinking,* or *watching the environment*. 97.3% of the participants would at least “sometimes” *watch the environment*, 49.0% would do so “very often”. With less than 40% of occasional (“sometimes”) use, *office work, watching a movie* and *gaming on a smartphone or tablet* are less popular than the aforementioned six NDRAs.

In Figure 11.2, the NDRAs are ordered by their popularity among the participants. The answers were given values from *0 = never* to *5 = very often*. The average scores of the NDRAs define the ranking in the figure:

1. watching environment (*M* = 4.30, *SD* = .806)
2. drinking (*M* = 3.64, *SD* = .993)
3. texting (*M* = 3.44, *SD* = 1.309)
4. phone call (*M* = 3.25, *SD* = 1.323)
5. eating (*M* = 3.20, *SD* = 1.033)
6. app usage (*M* = 3.10, *SD* = 1.440)
7. office work (*M* = 2.03, *SD* = 1.541)
8. watching a movie (*M* = 1.64, *SD* = 1.503)
9. gaming on smartphone/tablet (*M* = 1.62, *SD* = 1.615)

A 2 x 9 factorial ANOVA with repeated measures on the second factor showed a significant main effect for the factor *NDRA* ($F(6.194, 891.911) = 111.049$, $p < .001$, $\eta_p^2 = .435$) which measured the frequency of engaging in the NDRAs. Furthermore, a significant interaction of both factors was found ($F(6.194, 891.911) = 2.501$, $p < .05$, $\eta_p^2 = .017$). A main effect of the factor *gender* was not found. The degrees of freedom were corrected with Greenhouse-Geisser and $\varepsilon = .774$. One person with diverse gender had to be excluded from this analysis due to the small sample size.

In post-hoc comparisons, the following differences were found (Bonferroni-corrected alpha error of .005):

- *watching the environment* was significantly more popular than all other NDRAs (for all $p < .001$)
- *drinking* was significantly more popular than all other NDRAs ($p < .001$), except *watching the environment*
- *texting* was significantly more popular than *watching a movie*, *office work* and *gaming* ($p < .001$)
- *phone calls*, *app usage* and *eating* were each significantly more popular than *watching a movie*, *office work* and *gaming* ($p < .001$)

The following correlations were found (please note: The participants' age and the years of them holding a driver's licence correlate with $r = .97$):

- the younger the participants were, the more likely they were willing to perform an NDRA ($r = -.360$, $p < .001$; exceptions: for *watching the environment*, *eating* and *phone calls* no statistically significant relationship was found)
- the less driving experience (years of holding a driver's licence) the participants had, the more likely they were willing to perform an NDRA ($r = -.321$, $p < .001$; exceptions: for *watching the environment*, *eating* and *phone calls* no statistically significant relationship was found)
- the higher the pre-existing knowledge about automated driving of the participants was, the more likely they were willing to engage in an NDRA ($r = .315$, $p < .001$; individual analyses significant for *texting*, *watching a movie*, *office work* and *gaming*)
- the higher usefulness of automated driving was rated before the presentation of the nine NDRAs, the more likely the participants were willing to engage in NDRAs ($r = .308$, $p < .001$, individual analyses significant for *texting*, *phone calls, watching movies, office work* and *gaming*)
- the higher satisfaction with automated driving was rated before the presentation of the nine NDRAs, the more likely the participants were willing to engage in NDRAs ($r = .372$, $p < .001$, individual analyses significant for *texting, phone calls, app usage, watching movies, office work* and *gaming*)
- annual mileage did not seem to be a sufficient predictor for NDRA use ($r = .112$, $p = .177$)

In Level 3 automated vehicles, drivers have to take over vehicle control when the function issues a TOR. Thus, NDRAs have to be interrupted. For each NDRA, the participants were asked how they would assess an interruption by a TOR. For this assessment, the satisfaction scale of van der Laan's acceptance questionnaire and four additional word pairs were used. "Satisfaction" in this context is an indicator that shows how disturbing a TOR is perceived in the given situation.

Please note: In van der Laan's questionnaire, negative numeric values stand for high acceptance and vice versa. For an easier readability of the results, all satisfaction and usefulness values are reported inverted so that positive numeric values depict high acceptance and vice versa.

The NDRA were ranked by applying averaged satisfaction scores. The averaged satisfaction scores range from -2 (= lowest satisfaction) to +2 (= highest satisfaction). All but one NDRA were rated at least slightly below zero and thus show rather low satisfaction. Only "watching the environment" was rated as a moderately satisfying NDRA. The standard deviations were large, which indicates that the participants rated satisfaction with an interruption non-uniformly:

1. watching the environment (*M* = .469, *SD* = .896)
2. drinking (*M* = -.014, *SD* = 1.106)
3. app usage (*M* = -.026, *SD* = 1.008)
4. gaming on smartphone/tablet (*M* = -.073, *SD* = 1.119)
5. texting (*M* = -.104, *SD* = 1.015)
6. eating (*M* = -.138, *SD* = 1.013)
7. phone call (*M* = -.247, *SD* = 1.027)
8. watching a movie (*M* = -.361, *SD* = 1.175)
9. office work (*M* = -.493, *SD* = 1.182)

A 2 x 9 factorial ANOVA with repeated measures on the second factor showed a significant main effect for the factor *NDRA* ($F(5.918, 852.211) = 26.238$, $p < .001$, $\eta_p^2 = .154$) which measured the frequency of engaging in the NDRAs. Neither a main effect of the between-group factor *gender*, nor significant interactions was found. The degrees of freedom were corrected with Greenhouse-Geisser and $\varepsilon = .740$. One person with diverse gender had to be excluded from this analysis due to the small sample size.

In post-hoc comparisons, the following differences were found (Bonferroni-corrected alpha error of .005):

- satisfaction with being interrupted while *watching the environment* was significantly higher than for all other NDRAs (for all $p < .001$)
- satisfaction with being interrupted while *drinking* was significantly higher than for *office work* ($p < .001$)
- satisfaction with being interrupted while *using apps* was significantly higher than for *watching movies* ($p < .001$) and *office work* ($p < .001$)
- satisfaction with being interrupted while *gaming on smartphone/tablet* was significantly higher than for *office work* ($p < .001$)
- satisfaction with being interrupted while *texting* was significantly higher than for *office work* ($p < .001$)

- satisfaction with being interrupted while *eating* was significantly higher than for *office work* ($p < .001$)

The following correlations were found:

- the younger the participants were, the lower the satisfaction with being interrupted by a TOR while being engaged in a NDRA was (r =.236, p = .004, exceptions: for *watching a movie* and *watching the environment* no statistically significant relationship was found)
- the higher the driving experience (in years) was, the higher satisfaction with being interrupted during an NDRA was (r =.231, p = .005, exceptions: for *watching a movie* and *watching the environment* no statistically significant relationship was found)
- the higher the annual mileage was, the higher is satisfaction with being interrupted during a NDRA was (r =.280, p = .001, exception: for *watching the environment* no statistically significant relationship was found)
- pre-existing knowledge on automated driving did not seem to be a sufficient predictor for satisfaction with NDRA interruption (r = -.026, p = .751)

To gain deeper insights into the participants' perception of interruptions of NDRAs during automated drives, four additional items, taken from Arndt (2011), were used for assessing the NDRA and presented along with the satisfaction scale. The word pairs of the items represent the negative and positive poles on a five step Likert scale (-2 to +2):

- uncomfortable (-2) … comfortable (+2)
- unimportant (-2) … important (+2)
- incontrollable (-2) … controllable (+2)
- dangerous (-2) … safe (+2)

The following figures show the results of a descriptive analysis for each word pair.

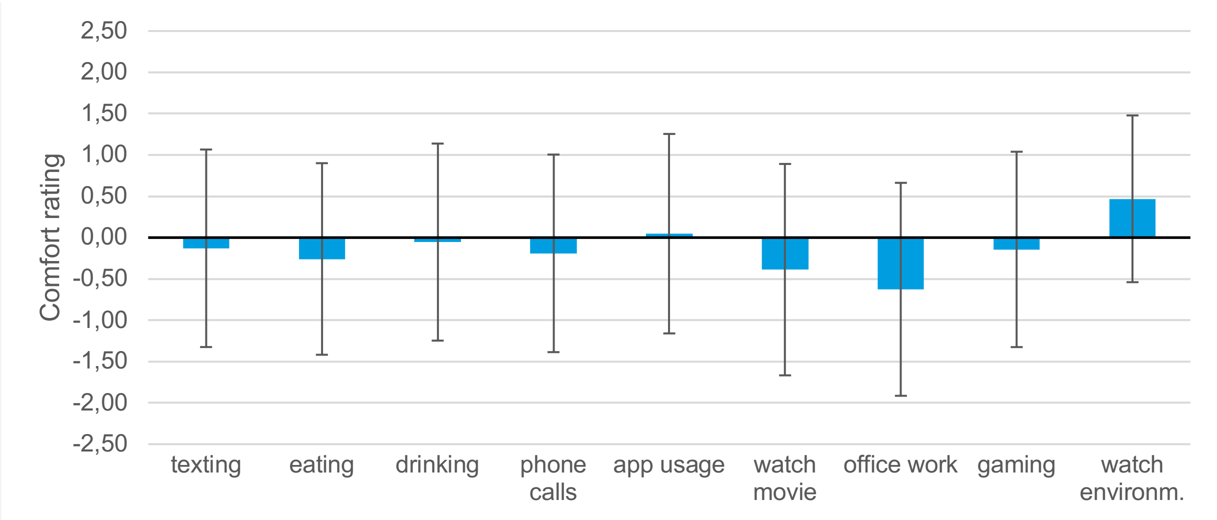

Figure 11.3: "Being interrupted while [NDRA] would be... comfortable" (+2) / uncomfortable (-2), means with standard deviations

For most of the presented NDRAs, an interruption by a TOR was – on average – perceived as rather neutral in terms of comfort, with minor amplitudes to the negative ("uncomfortable") or positive ("comfortable") side. The high standard deviations for all NDRAs indicated that the participants rated comfort non-uniformly. Interruptions of *office work* and *watching a movie* were perceived as most uncomfortable by the participants. With a mean of .47, having to stop to *watch the environment* was rated most comfortable of all NDRAs.

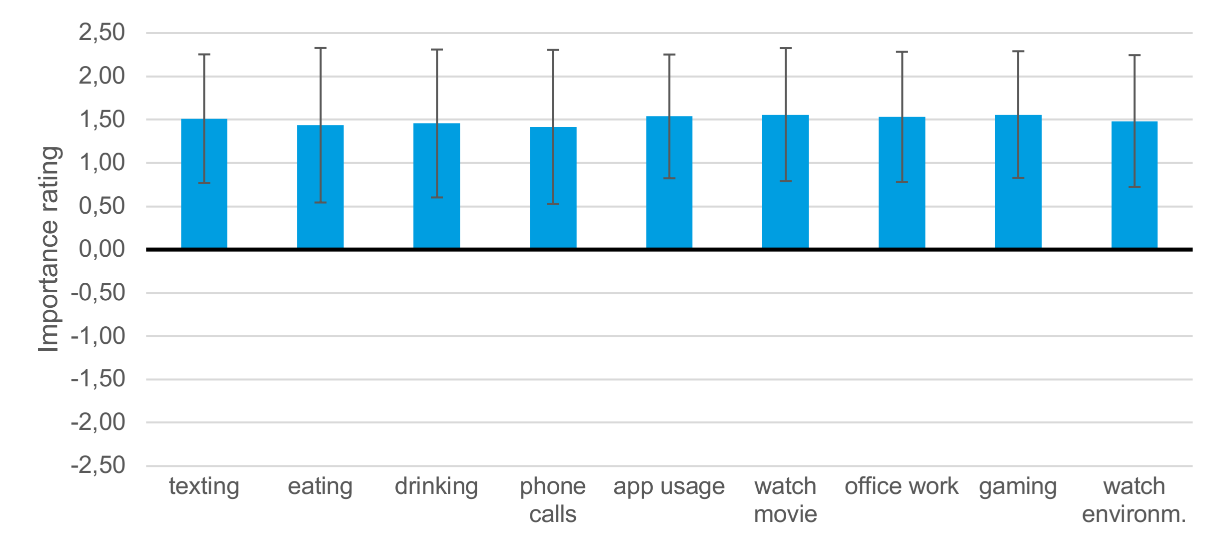

Figure 11.4: "Being interrupted while [NDRA] would be... important" (+2) / unimportant (-2)", means with standard deviations

On average, the participants perceived an interruption of an NDRA as important with only minor differences between the NDRAs (means between 1.41 and 1.56). The standard deviations for all

NDRAs were high, which means that the participants rated the importance of an interruption non-uniformly.

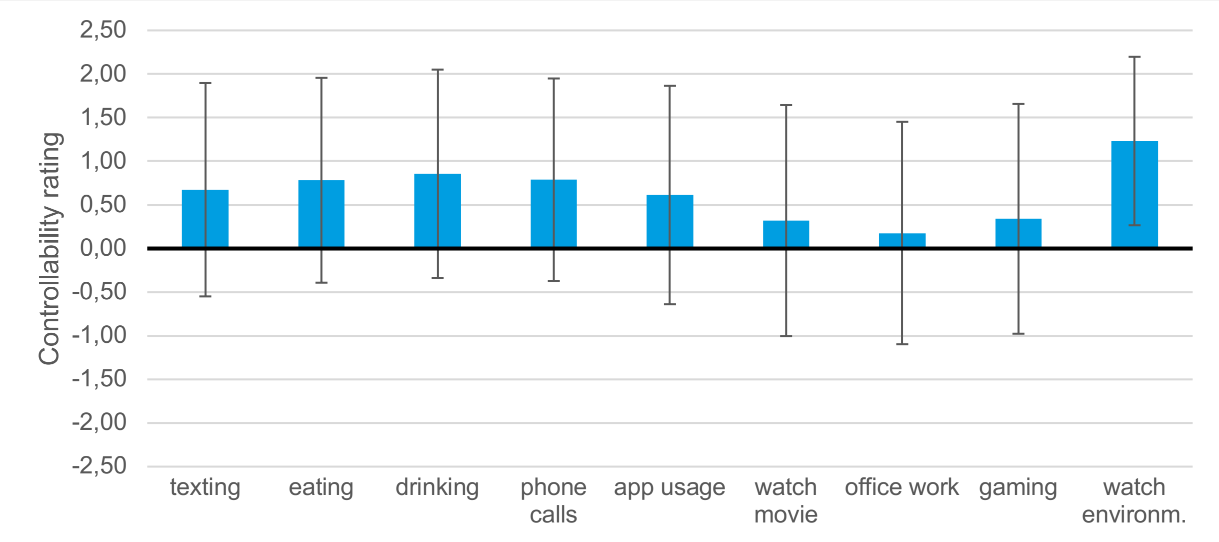

Figure 11.5: "Being interrupted while [NDRA] would be... controllable" (+2) / uncontrollable (-2)"; means with standard deviations

For all of the nine NDRAs, an interruption was on average rated in the positive, thus "controllable" range. Stopping to *watch the environment* was assessed as most controllable, performing *office work* as least controllable. The high standard deviations for all NDRAs indicate that the participants rated the controllability of an interruption non-uniformly.

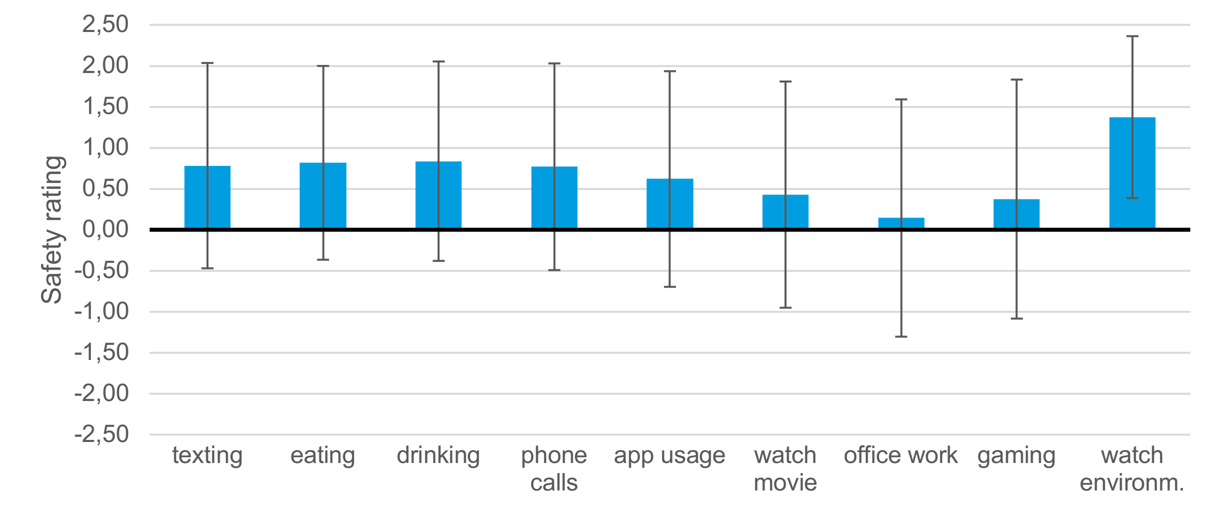

Figure 11.6: "Being interrupted while [NDRA] would be... safe" (+2) / dangerous (-2)", means with standard deviations

Overall, the participants perceived an interruption as neutral to positive in terms of safety. Larger differences were seen nonetheless: With a mean of 1.37, *watching the environment* was assessed safest, while *office work* (0.14) was rated close to neutral and thus worst. The standard deviations for all NDRAs were high, which means that the participants rated the safety of an interruption non-uniformly.

11.4 Summary

In the present online study, 147 participants were presented videos on automated driving and asked about their general acceptance towards Level 3 automation and their perception of NDRAs during automated driving.

In terms of general acceptance towards automated driving, both usefulness and satisfaction were fairly high rated by the participants. Males found automated driving significantly more useful and satisfying than females, significant differences among age groups could not be found.

The most popular NDRAs were watching the environment, as well as many (smart)phone related NDRAs (texting, phone calls, app usage) and food consumption (eating & drinking). Office work, watching a movie and gaming on smartphone/tablet were less popular. Younger participants and participants with higher preknowledge about automated driving were more likely to engage in NDRAs and showed a higher acceptance towards automated driving in general.

On average, participants perceived an interruption of an NDRA due to a TOR as neutral to dissatisfying, but ratings differ widely among the participants. *Watching the environment* was the only NDRA of which an interruption was perceived as modestly satisfying. The younger the participants were, the lower their satisfaction was. Importance of interruptions by TOR received high ratings. This view of the participants varied only slightly between NDRAs. In terms of safety and controllability, interruptions of NDRAs by TORs tended to receive positive ratings by the participants.

11.5 Discussion of Results

The overall acceptance of automated driving was fairly high in this online study. The presentation of the nine videos showing NDRAs and their interruption by TORs did not affect participants' acceptance in a before-after-comparison: The role of the driver/user in a Level-3-ADF was explained to the participants before the first assessment of acceptance and could already have established a realistic expectation towards automated driving that was not further influenced by the videos.

Nearly all participants would at least sometimes *watch the environment* during automated drives: This result seems plausible since this NDRA does not need any further equipment and after activating the ADF, users are watching the environment by default. The high popularity of *watching the environment* could also mean that users do not always want to be distracted or kept busy during automated driving. However, recent research shows that prolonged periods of *watching the environment* in Level 3 (e.g., due to daydreaming) can lead to increased driver fatigue and vigilance decrement (Frey, 2021).The use of (smart)phones for *texting, phone calls* and *app usage* was also attractive to the participants of this online study. The high availability of smartphones and the numerous activities linked to them probably affected this result. Furthermore, food consumption as a basic human need was a favoured NDRA. It could be advisable to check whether smartphone use and food consumption can be made possible in production automated vehicles in order to meet potential users' requirements towards NDRA choice. *Office work, watching a movie* and *gaming on smartphones/tablets* were less popular: It depends on the user's profession whether office work is even a necessity or possibility in general (let alone in a car). In addition, it is a matter of personal taste whether someone likes movies or gaming and if a car is perceived as a good place to enjoy these.

Interrupting a NDRA due to a TOR was – to a certain extend – dissatisfying for the participants, with the exception of *watching the environment*. This could mean that NDRA engagement is rather important to potential users. Younger participants are more likely to engage in an NDRA and are also more dissatisfied with interruptions: This can emphasise the importance of investigating NDRA engagement and overall perception of automated driving since future customers and users seem to have differing expectations.

For all NDRAs, an interruption was perceived as important by the participants: That might reflect a correct understanding of the system and the user's role. It is also perceived as rather safe and controllable; that could mean that the participants think a safe take-over process is possible with the respective NDRA, at least in theory.

Overall, users do accept ADFs, but the medium to negative satisfaction with interruptions of NDRAs needs further investigation. TORs are an inevitable part of Level 3 automation, and the results have shown that users are willing to engage in different NDRAs: Especially for younger users, conflicts between the necessity to take-over and the engagement in NDRAs can arise. It should be carefully assessed how the users' needs and their role as a fallback for the system can be balanced.

12 Summary of results

12.1 Long-term behavioural adaptation

The following table provides a summary of the results on long-term behavioural adaptation. As can be seen, results on behavioural adaption are mixed across studies. Especially the on-road study (see chapter 6) which took place on public roads with a prototype ADF found no changes in driver behaviour and acceptance with repeated usage. Here, like for some measures used in the simulator study (see chapter 5), the main reason for this was presumably a ceiling effect with already very positive perceptions stated during the first contact with the ADF. Results from simulator study and case study (see chapter 4) indicate that with growing experience, the evaluation of the ADF becomes more positive with increasing trust. There is little room for measurable changes on indicators due to ceiling effects after 1st contact. In none of the studies and for none of the investigated indicators was there a decrease of acceptance with repeated usage.

In summary, it seems that the perception of the ADF becomes more positive with growing experience or remains on the same, mostly high level. On objective parameters of system handling, this is reflected in an increase of time spent on side activities and in a decrease of time spent on supervising the ADF and the surrounding traffic environment.

Table 12.1: Summary of results on long-term behavioural adaptation. Green indicates that the hypothesis is clearly supported by results, grey that there are no significant results and yellow that results are mixed. Non-signifcant results are counted as not supporting the hypothesis. There were no results that explicitely contradicted the hypotheses which would be marked in red.

ID	Specific hypotheses	Case study BA	Simulator BA	Pilot study BA
RQ-U1	Willingness to use increases with increasing experience with function.			
RQ-U3	Perceived safety increases with increasing experience with function.			
	Perceived comfort increases with increasing experience with function.			
	Perceived reliability increases with increasing experience with function.			
	Trust increases with increasing experience with function.			
RQ-U4	With increasing experience, understanding of the system increases.			
RQ-U5	Over AD usage time, drivers experience less stress.			

ID	Specific hypotheses	Case study BA	Simulator BA	Pilot study BA
	After a familiarisation period, drivers will become drowsy more rapidly.			
	Over AD usage time, drivers experience less workload.			
RQ-U6	With increasing experience, attention to other road users decreases.			
RQ-U9	Secondary task interaction increases with increasing experience with function.			
RQ-U10	Take-over performance increases with increasing experience with function.			
RQ-U11	Pattern of system activation will become more dependent on driving scenario with increasing experience with function.			

Independent of the used methodological approach (simulator, Wizard of Oz, on-road study), results indicate that drivers were either highly positive about the tested ADF from the beginning or became more confident with repeated usage. However, this did not manifest in growing knowledge of the ADF or in drivers' increased adaptation to the situation.

The main effect of behavioural changes over time occurred after the first usage of the ADF. Trust, acceptance and the willingness to use the ADF increased after the first drive and then remained stable on a high level (see chapter 5).

12.2 Short-term behavioural adaptation

The potential impact of ADF use on immediate driving behaviour following a section of automated driving was studied in two ways:

1. The impact of ADF implementation on driving behaviour either during the transition to manual driving of after the transition to manual driving.
2. The impact of duration of driving with ADF active on the transition to manual driving.

Results showed that the use of ADFs had an immediate effect on different driver behaviour and acceptance. This was especially evident in, but not limited to, take-over situations.

Impact on driver behaviour in take-over situations was investigated in several studies. In wizard of Oz studies (see chapter 7), immediate take-over performance and reaction times were independent of the duration of driving with the ADF engaged directly before the take-over request and the timing of the take-over requests. Adaptation of drivers' reaction to the timing of a take-over request was found with growing experience of an ADF over several drives and multiple experienced take-over situations (see chapter 5). Overall, it can be concluded that drivers tended to react quickly to a take-over request during first contact. It requires repeated experience of take-over situations to learn to use a time budget provided by the ADF. In the simulator study, reaction times increased significantly with growing experience when the ADF allowed drivers to take their time.

In the ADEST study (see chapter 7) about one third of drivers responded late and therefore crashed with a stationary on-road object not sensed by automation, while supervising a near-perfect L2 automation system. A hands-on-wheel requirement did not influence when drivers started to steer to pass the object nor if drivers crashed or not. High trust in automation was associated with delayed response and crashing.

In the TJP study, drivers' take-over performance (take-over times and driving performance in a road-work zone) was found to not be considerably influenced by automation duration (4.5 minutes vs. 14 minutes). In fact, the effect of automation (L3 vs. manual) was greater than the effect of automation duration. Drivers started to steer earlier (farther away) away from a road-works zone after L3 automation (both durations) compared to the manual driving baseline.

In the L3Pilot test track study, drivers take-over performance was not influenced by the timings of the take-over requests in response to the conflict object (i.e. the take-over time budget). That is, drivers used a similar amount of time for their first glance to the instrument cluster, putting hands on wheel, glancing to the forward road and deactivating automation. However, issuing the take-over request early may result in drivers that slow down before a conflict object becomes visible (precautionary braking) and that have more time to assess the status of the automated system and the surrounding environment before stabilizing the gaze to the forward road.

In the L3Pilot Woz pilot all drivers managed to resume manual driving from automation in response to the take-over request. The longest observed take-over time was about 9 s. All drivers eventually placed their feet on the accelerator pedal, whereas only a few drivers touched the brake pedal.

Besides impact on behaviour during the transition of control, effects of driving with an ADF on continuous manual driving afterwards was in the focus of interest. Manual driving behaviour was affected by the experience of automated driving (chapter 9). When drivers were exposed to a short time distance to the lead vehicle (THW) during automated driving, they also chose short THWs during manual driving. This finding represents behavioural adaptation in the classical sense in that behavioural changes "which were not intended by the initiators of the change" (OECD, 1990) is observed. The focus of this study was a comparison between the two different THW settings and not a comparison to the baseline behaviour, which is reported mainly for transparency. This simulator study was conducted to initiate a scientific discussion, and the short THW setting was chosen to investigate an extreme automated driving configuration. It is not likely that such a setting will be implemented in the near future. This study is limited to the investigation of short-term effects. Hence, it only allows for very limited conclusions on the effect of automated driving on the general car following behaviour.

12.3 Impact of ADF level on driver behaviour and system evaluation

Drivers rated an L4 motorway chauffeur more positively than an L3 motorway chauffeur although the two implementations were tested in a between-group design (see chapter 5). Furthermore, there was no impact of ADF-level on the development of fatigue, neither in the comparison L3 vs. L4 (see chapter 5) nor in the comparision L2 vs. L3 (see chapter 8). Across studies, there was a slight increase of reported fatigue of 0.5 points on KSS-scale during a 30-minute drive.

The average change of KSS as a function of drive time was remarkably similar in different levels of automation (L0-L3) and alcohol (BAC = 0.0 or 0.1%, see chapter 8). A general trend indicated a larger standard deviation of KSS change during the drive in L3 compared to L1 (L3pilot ASTA study), and while intoxicated in L3 compared to the L3 baseline and lower levels of automation.

Drivers directed considerably less attention to the road in L3 automation compared to manual driving (L0) and L2 automation in the sober baseline drive of the impairment study. The effect of alcohol (BAC 0.1%) increased the PRC during non-task segments, while decreased PRC during secondary tasks was observed for all levels of automation compared to baseline.

During secondary tasks, the off-path glance durations were considerably longer in L3, compared to L0-L2. The effect of alcohol further amplified the effect of automation on the long off-path glances during secondary tasks.

12.4 Non-driving related activities while driving with ADF active

In the Wizard of Oz case study (see chapter 4), participants mainly used their smartphone or read a magazine when engaging in non-driving related tasks during phases of automated driving. Other tasks such as the use of a tablet PC or office work were observed less frequently. The time spent with NDRAs varied widely: For three of six participants, an increase in NDRA engagement over time was observed. Two participants spent the vast majority of automated driving time in all three drives with NDRAs so that there was barely room for a further increase. One participant spent only little time with NDRAs during his first and third drive, but more than half of the time in the second drive.

With increasing experience with an ADF, drivers spent an increasing amount of time driving with the ADF on non-driving related tasks (see chapter 5). Also in the simulator study, drivers mainly engaged with their smartphone. However, engagement in other activies, like reading, eating or drinking or doing paper work was also observed. This preference is supported by questionnaire data collected fom the participants at the end of the experimental sessions. Drivers who had experienced a L4-ADF stated that they would watch movies and sleep more frequently than drivers who had used a L3-ADF.

The online study on user acceptance and NDRA engagement (see chapter 11) supports these findings and shows that the most popular NDRAs were “watching the environment”, as well as many (smart)phone-related NDRAs (texting, phone calls, app usage) and food consumption (eating & drinking). Office work, watching a movie and gaming on smartphone/tablet were less popular.

12.5 Take-over situations

Transitions of control or take-over situations were one of the main focuses of the supplementary studies. Wizard of Oz and driving simulator studies as well as online surveys focused on driver acceptance of and behaviour during take-over situations.

The criticality of a take-over situation generally depends on a variety of factors, among others, the traffic situation, e.g., presence of other vehicles or conflict objects, the driver state and the take-over modality. The frequency of crashes was not affected by the requirement to put the hands on the wheel during AD (chapter 7.3). However, the drivers' trust level affected the crash rate: High-trust drivers crashed more frequently than low trust drivers.

The results on the effects of take-over time budget on the drivers' take-over response was mixed: Drivers provided with a large time budget of 45s showed later responses to a TOR than drivers provided with a take-over time budget of 15s in a driving simulator study (chapter 5). In a Wizard of Oz test track study, there was no effect of time budget (9s vs. 18s) on the take-over response time (chapter 7.3). It has to be considered that after increasing experience with the ADF and take-over situations, drivers' take-over reponses were prolonged in the 45s condition of the simulator study (chapter 5). In the Wizard of Oz study, the take-over situation was only presented once and therefore, no potential changes in take-over response were captured.

When a short take-over time was provided, there was no change of take-over response with repeated experience of TORs (see chapters 5 and 10). However, with a longer take-over time budget, drivers' take-over responses were delayed after repeated experience of TORs (chapter 5). This delayed reaction was not associated with a decreased take-over performance.

The take-over modality had no effect on the take-over response or trust and perceived safety. However, drivers preferred an auditory warning over a peripheral light (chapter 10).

12.6 The impact of driver state on acceptance and usage of ADFs

Acceptance and usage of ADF was not only affected by conditions of the system but also by conditions of the driver. Drivers reported less stress and lower workload with repeated usage. They also directed less visual attention to the road with increasing trust levels. Despite some empirical evidence from other research (as discussed in section 3.4), drivers did not agree with the statement that automated driving would make them tired. The drivers' sleepiness did not increase during automated driving (chapter 5), not even when drivers had a BAC level of 0.1% (chapter 8). However, the drivers' visual attention was affected by alcohol intoxication: When intoxicated, drivers directed more attention to the road when they were not engaged in a secondary task and less attention when they were engaged in a secondary task.

Even though automated driving did not increase driver sleepiness, the drivers' behaviour changed significantly when they were sleepy: Unsurprisingly, the drivers closed their eyes for longer periods of time and a significant part of drivers fell asleep (section 5.5.11). However, the drivers' evaluation of the ADFs was not affected by sleepiness.

13 Conclusions and recommendations for practice

Within L3Pilot, a variety of user-related topics was addressed in the supplementary studies. For this purpose, different methodological approaches were used. The findings of the supplementary studies helped to create the “bigger picture” of user behaviour and acceptance of AD and changes over time.

13.1 Behavioural changes and safety

Behavioural adaptations to ADFs can potentially have consequences for overall safety. Increasing trust with increasing usage can generally be seen as a positive development. However, increasing trust is accompanied with increasing engagement in tasks that involve both hands, and thus might compromise take-over performance, lead to longer take-over times and even a misuse of the system by sleeping (chapter 5). Driver monitoring systems should be able to detect any adverse behaviour and the ADF needs to take action to stop or prevent that behaviour.

The design of the system or HMI has also proven to affect drivers’ behaviour and should be designed to account for behavioural adaptations. Drivers reduced their time distance to the lead vehicle during manual driving when they had experienced shorter distances during automated driving. This finding emphasises that behavioural adaptations go beyond driver behaviour during ADF usage and shape driver behaviour during manual driving, too. Designers of ADFs should take into account that the drivers’ behaviour can be affected on many levels by ADF use and even during manual driving.

With higher levels of automation, drivers spend less attention on the road and monitor the driving environment less. As a consequence, drivers might be “out of the loop”.

13.2 Recommendations for future research

Regarding the fundamental changes that the introduction of L3-ADFs will bring for the tasks and responsibilities of drivers, the topic of behavioural adaption will remain a relevant one. Based on the supplementary studies in L3Pilot some methodological conclusions for future research can be drawn:

- Relevant behavioural changes can be observed already after two or three times of system usage. Therefore, for many purposes medium term setups with a few measurement points but not usage over weeks might be a good starting point.
- Some behavioural adaptions to specific decision in system / HMI-design (e.g. transition times) become more pronounced with repeated usage of a system. To study those, experimental setups focusing on first contact with an ADF might be not the best approach.
- The ceiling effect especially in the on-road study on long-term behavioural adaption highlights that indicators for measuring behavioural changes should be chosen carefully. They need to give room for changes. Indicators, e.g. questionnaire items, that tend to elicit highly positive

ratings already after first contact with an ADF are not the best choice because with them behavioural changes might be not measurable.

13.3 Recommendations for practice

The greatest behavioural adaptation was evident between the first and the second drive. For future studies on driver behaviour in AD, this could mean that in order to make predictions about the users' actual behaviour, at least two test sessions should take place. However, further relevant changes in behaviour might be observable over a longer time frame. It has to be considered that all presented studies followed a more or less standardised protocol and drivers did not show "naturalistic" behaviour. For instance, they were not free to use an ADF at any time or on any road. More naturalistic testing approaches could give deeper insights into driver behaviour and changes in driver behaviour.

The studies indicate that L3-ADFs might be prone to misuse by drivers. In the simulator study but also in a pilot study on public roads, drivers indicated that they would use such a system at least now and then to sleep while driving. In the simulator study it became clear that drivers knew that they were not allowed to sleep but to remain attentive all the time. Still, they felt it was safe to sleep because they were still able to handle the take-over situations. This is in line with results reported for the pilots (see Weber et al. 2021). This indicates that driver monitoring systems might be necessary in order to prevent misuse.

Hand in hand with the issue of potential misuse goes the problem of overtrust in an ADF. In a test track study, it could be shown that drivers who reported higher levels of trust in the system were more frequently not able to handle a take-over situation safely. For the introduction of L3 ADFs it might be helpful to develop strategies that support drivers to develop a realistic picture of the capabilities of ADFs over time. This could help to prevent overtrust and maybe also misuse of an ADF.

In several studies, it turned out that drivers used the time while driving with an ADF active for other none driving-related tasks. Interaction with a smartphone was one of the most popular activities. Regarding safety, this is first of all good news because smartphones are rather small devices that allow the driver to still look towards the road if necessary and to have at least one hand free for reactions required for driving. However, it might further enhance safety if smartphones and their applications could be designed in a way that tasks can be interrupted more easily if necessary. Here, new solutions are required that help to integrate (popular) side tasks and the HMI of the ADF to keep the driver in the loop and to support safe transitions of control.

List of publications

Gonçalves, R., Lyu, W., Torrão, G., Puente Guillen, P., Louw, T., & Merat, N. (2020, February). Development of an algorithm to identify stabilisation time for car-following after transitions of control from vehicle automation. 12th International Conference on Methods and Techniques in Behavioral Research and 6th Seminar on Behavioral *Methods Volume 1 of the Proceedings of the joint meeting of the 12th International Conference on Measuring Behavior and the 6th Seminar on Behavioral Methods* to be held in Krakow, Poland, October 15-18, 2021.

Louw, T., Goncalves, R., Torrao, G., Radhakrishnan, V., Lyu, W., Guillen, P. P., & Merat, N. (2020). Do drivers change their manual car-following behaviour after automated car-following? *Cognition, Technology & Work*, 1-15.

Louw, T., Madigan, R., Lee, Y.M., de Marco, C., Mallada, J.L., Merat, N. (2021). Don't Worry, I'm in Control! Is Users' Trust in Automated Driving Different When Using a Continuous Ambient Light HMI Compared to an Auditory HMI? In *Proceedings of the 13th International Conference on Automtove User Interfaces and Interactive Vehicular Applications*.

Lyu, W., Gonçalves, R., Guo, F., Torrão, G., Radhakrishnan, V., Puente Guillen, P., Louw, T., & Merat, N. (2021). Applying Entropy to Understand Drivers' Uncertainty during Car-following. 12th International Conference on Methods and Techniques in Behavioral Research and 6th Seminar on Behavioral Methods *Volume 1 of the Proceedings of the joint meeting of the 12th International Conference on Measuring Behavior and the 6th Seminar on Behavioral Methods* to be held in Krakow, Poland, October 15-18, 2021.

Metz, B., Wörle, J., Hanig, M., Schmitt, M. & Lutz, A. (2020). Repeated Usage of an L3 Motorway Chauffeur: Change of Evaluation and Usage. *Information* (Switzerland) 11(2):114.

Metz, B., Wörle, J., Hanig, M., Schmitt, M. & Lutz, A. (2020). *Change of acceptance with repeated usage of a L3-motorway chauffeur.* Transport Research Arena 2020, Helsinki, Finland. (Conference canceled).

Metz, B., Wörle, J, Hanig. M. , Schmitt, M. Lutz, A. & Neukum, A. (2021). Repeated usage of a motorway automated driving function: Automation level and behavioural adaption. *Transportation Research Part F Traffic Psychology and Behaviour*, 81(4):82-100. DOI: 10.1016/j.trf.2021.05.017

Pipkorn, L., Victor, T. W., Dozza, M., & Tivesten, E. (2021a). Driver conflict response during supervised automation: Do hands on wheel matter?. *Transportation Research Part F: Traffic Psychology and Behaviour*, 76, 14-25. https://doi.org/10.1016/j.trf.2020.10.001

Pipkorn, L., Victor, T., Dozza, M., & Tivesten, E. (2021b). Automation Aftereffects: The Influence of Automation Duration, Test Track and Timings. *IEEE Transactions on Intelligent Transportation Systems*. Doi: 10.1109/TITS.2020.3048355

Pipkorn, L., Tivesten, E., & Dozza, M. (2021). *It's about time! Early take-over request in conditional automation enables safe response to a lead-vehicle cut-out scenario*. Manuscript submitted for publication.

Pipkorn, L., Dozza, M., & Tivesten, E. (2021). *Driver glance behaviour before and after take-over requests in conditional automation on public road*. Manuscript submitted for publication.

Radhakrishnan, V., Merat, N., Louw, T., Gonçalves, R. C., Lyu, W., Torrao, G., ... & Lenné, M. G. (2021). Physiological Indicators of Driver Workload During Car-Following Scenarios and Takeovers in Highly Automated Driving. *PsyArXiv*. DOI: 10.31234/osf.io/bwp6f

Tivesten, E., Broo, V, Ljung Aust, M. (2021). *The influence of alcohol, automation, and non-driving related tasks on driver visual behavior on test track*. Manuscript submitted for publication.

Wörle, J. & Metz, B. (2020). Driving with an L3 - motorway chauffeur: How do drivers use their driving time? In D. de Waard, A. Toffetti, L. Pietrantoni, T. Franke, J-F. Petiot, C. Dumas, A. Botzer, L. Onnasch, I. Milleville, and F. Mars (Eds.), *Proceedings of the Human Factors and Ergonomics Society Europe Chapter* 2019 Annual Conference.

Wörle J., Metz B., Lutz A. & Schmitt M. (2020). Automated Driving on the Motorway: A Users' Perspective on Conditional Versus High Automation. In Stanton, N. (Ed.), *Advances in Human Aspects of Transportation*. AHFE 2020. Advances in Intelligent Systems and Computing (pp 166-172). Cham: Springer.

Wörle, J., Metz, B. & Baumann, M. (2021). Sleep inertia in automated driving: Post-sleep take-over and driving performance. *Accident Analysis and Prevention*, 150, 105918.

Wörle, J., Metz, B., Steinborn, M. B., Huestegge, L., & Baumann, M. (2021). Differential effects of driver sleepiness and sleep inertia on driving behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, 82, 111-120.

References

AASM (2017). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Darien, IL: American Academy of Sleep Medicine;

Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior 1. *Journal of applied social psychology*, *32*(4), 665–683.

Akerstedt, T. and Gillberg, M. Subjective and objective sleepiness in the active individual. *Int. J. Neurosci.*, 1990, 52: 29–37

ASTAzero. (2021). [online]. Retrived on the 9th of June 2021. http://www.astazero.com/

Arndt, S. (2011). *Evaluierung der Akzeptanz von Fahrerassistenzsystemen*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Beggiato, M., & Krems, J. F. (2013). The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information. T*ransportation research part F: traffic psychology and behaviour*, 18, 47–57.

Bellem, H., Thiel, B., Schrauf, M., & Krems, J. F. (2018). Comfort in automated driving: An analysis of preferences for different automated driving styles and their dependence on personality traits. *Transportation research part F: traffic psychology and behaviour*, 55, 90–100.

Borojeni, S. S., Chuang, L., Heuten, W., & Boll, S. (2016, October). Assisting drivers with ambient take-over requests in highly automated driving. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 237-244).

de Waard, D., van der Hulst, M., Hoedemaeker, M., & Brookhuis, K. A. (1999). Driver behavior in an emergency situation in the Automated Highway System. *Transportation human factors,* 1(1), 67-82.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., ... & Bucher, C. (2006). *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment* (No. DOT-HS-810-593). United States. Department of Transportation. National Highway Traffic Safety Administration.

Dikmen, M., & Burns, C. (2017). *Trust in autonomous vehicles: The case of Tesla Autopilot and Summon*. Paper presented at the Systems, Man, and Cybernetics (SMC), Banff, Canada

Dufour, A. (2014). *Driving Assistance Technologies and Vigilance: Impact of Speed Limiters and Cruise Control on Drivers' Vigilance*. Seminar on the Impact of Distracted Driving and Sleepiness on Road Safety, April. Paris La Défense.

Dziennus, M., Kelsch, J., & Schieben, A. (2016). Ambient light based interaction concept for an integrative driver assistance system–a driving simulator study. *Proceedings of the Human Factors and Ergonomics Society Europe*, 171-182.

Eick, E. M., & Debus, G. (2005). Adaptation effects in an automated car following scenario. *Traffic and transport psychology: Theory and application*, 243-255.

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

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

Eriksson, A., & Stanton, N. A. (2017). Takeover Time in Highly Automated Vehicles: Noncritical Transitions to and From Manual Control. *Human Factors*, 59(4), 689–705. https://doi.org/10.1177/0018720816685832

FOT-Net (2018). *FESTA Handbook*. Version 7. Retrieved from https://connectedautomateddriving.eu/wp-content/uploads/2019/01/FESTA-Handbook-Version-7.pdf.

French, D. J., West, R. J., Elander, J., & Wilding, J. M. (1993). Decision-making style, driving style, and self-reported involvement in road traffic accidents. *Ergonomics*, 36(6), 627-644.

Frey, A. T. (2021). *Zum Fahrerzustand beim automatisierten Fahren: Objektive Messung von Müdigkeit und ihre Einflussfaktoren* (Dissertation, Technische Universität Braunschweig). https://doi.org/10.24355/dbbs.084-202103171057-0

Friedrich, B. (2016). The effect of autonomous vehicles on traffic. In *Autonomous Driving* (pp. 317-334). Springer, Berlin, Heidelberg.

Gipps, P. G. (1981). Behavioral car-following model for computer simulation. *Transportation Research Part B: Methodological*, 15(2), 105-111.

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). “Take over!” How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1), 1938–1942. https://doi.org/10.1177/1541931213571433

Gold, C., Körber, M., Hohenberger, C., Lechner, D., & Bengler, K. (2015). Trust in automation– Before and after the experience of take-over scenarios in a highly automated vehicle. *Procedia Manufacturing*, 3, 3025–3032.

Hartwich, F., Beggiato, M., & Krems, J. F. (2018). Driving comfort, enjoyment and acceptance of automated driving – effects of drivers’ age and driving style familiarity. *Ergonomics*, 1–16.

Heikoop, D. D., de Winter, J. C., van Arem, B., & Stanton, N. A. (2019). Acclimatising to automation: Driver workload and stress during partially automated car following in real traffic. *Transportation Research Part F: Traffic Psychology and Behaviour*, 65, 503-517.

Herman, R., Montroll, E. W., Potts, R. B., & Rothery, R. W. (1959). Traffic Dynamics: Analysis of Stability in Car Following. *Operations Research*, 7(1), 86–106. https://doi.org/10.1287/opre.7.1.86

Herzberger, N. D., Eckstein, L., & Schwalm, M. (2018). *Detection of Missing Takeover Capability by Orientation Reaction to a Takeover Request.* Paper presented at the 27th Aachen Colloquium Automobile and Engine Technology 2018, Aachen.

Itkonen, T. H., & Lehtonen, E. (2020). Characterisation of motorway driving style using naturalistic driving data. *Transportation Research Part F: Traffic Psychology and Behaviour*, 69, 72-79.

Jamson, A. H., & Merat, N. (2005). Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(2), 79-96.

Jamson, H., Merat, N., Carsten, O. M., & Lai, F. C. (2013). Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation Research Part C: Emerging Technologies*, 30, 116–125.

Jamson, S. L. (2013). The Role of Acceptance in Behavioural Adaptation. Behavioural adaptation and road safety: *Theory, evidence and action*, 289.

Karrer, K., Glaser, C., Clemens, C., & Bruder, C. (2009). Technikaffinität erfassen–der Fragebogen TA-EG. *Der Mensch im Mittelpunkt technischer Systeme*, 8, 196-201.

Körber, M. (2018, August). Theoretical considerations and development of a questionnaire to measure trust in automation. In *Congress of the International Ergonomics Association* (pp. 13-30). Springer, Cham.

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied ergonomics*, 66, 18–31.

Kreuzmair, C., Gold, C., & Meyer, M.-L. (2017). *The influence of driver fatigue on take-over performance in highly automated vehicles.* Paper presented at the 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Detroit, Michigan.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25(1), 178-206.

Kundinger, T., Wintersberger, P., & Riener, A. (2019). (Over) Trust in Automated Driving: The Sleeping Pill of Tomorrow? Paper presented at the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems.

Kunze, A., Summerskill, S. J., Marshall, R., & Filtness, A. J. (2019, September). Conveying uncertainties using peripheral awareness displays in the context of automated driving. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 329-341).

Kyriakidis, M., de Winter, J. C., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., Martens, M.H., Bengler, K., Andersson, J., Merat, N., Reed, N., Flament, M., Hagenzieker, M. & Happee, R. (2017). A human factors perspective on automated driving. *Theoretical Issues in Ergonomics Science*, 1–27.

Kyriakidis, M., Happee, R., & de Winter, J. C. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation research part F: traffic psychology and behaviour*, 32, 127–140.

Large, D. R., Burnett, G., Morris, A., Muthumani, A., & Matthias, R. (2017). *A Longitudinal Simulator Study to Explore Drivers' Behaviour During Highly-Automated Driving*. Paper presented at the International Conference on Applied Human Factors and Ergonomics.

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

Hibberd et al. (2018). *Deliverable D3.1 From research questions to logging requirements.* https://l3pilot.eu/fileadmin/user_upload/Downloads/Deliverables/L3Pilot-SP3-D3.1-From_research_questions_to_logging_requirements-v2.0_DRAFT_for_website.pdf

Lee, J. D., Fiorentino, D., Reyes, M. L., Brown, T. L., Ahmad, O., Fell, J., Ward, N., & Dufour, R. (2010). *Assessing the feasibility of vehicle-based sensors to detect alcohol impairment* (Report No. DOT HS 811 358). August.

Lee, S. E., Llaneras, E., Klauer, S.G., and Sudweeks, J. (2007). *Analyses of Rear-End Crashes and Near-Crashes in the 100-Car Naturalistic Driving Study to Support Rear-Signaling Countermeasure Development*. (Report number DOT HS 810 846). Washington, DC: U.S. Department of Transportation.

Lyu, N., Deng, C., Xie, L., Wu, C., & Duan, Z. (2019). A field operational test in China: Exploring the effect of an advanced driver assistance system on driving performance and braking behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, 65, 730-747.

Martens, M. H., & Jenssen, G. D. (2012). Behavioral adaptation and acceptance. *Handbook of Intelligent Vehicles* (pp. 117–138): Springer.

Meschtscherjakov, A., Döttlinger, C., Rödel, C., & Tscheligi, M. (2015, September). ChaseLight: ambient LED stripes to control driving speed. In *proceedings of the 7th international conference on automotive user interfaces and interactive vehicular applications* (pp. 212-219).

Metz, B., et al.(2020): D3.3. Evaluation methods. L3 Pilot Deliverable L3 Pilot.

Merat, N., Jamson, A. H., Lai, F. C., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human factors*, 54(5), 762–771.

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., . . . Itoh, M. (2018). The “out-of-the-loop” concept in automated driving: Proposed definition, measures and implications. Cognition, Technology & Work, 1-12.

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27(5-6), 527–539.

Naujoks, F., Mai, C., & Neukum, A. (2014). The effect of urgency of take-over requests during highly automated driving under distraction conditions. *Advances in Human Aspects of Transportation*, 7(Part I), 431.

Naujoks, F., & Totzke, I. (2014). Behavioral adaptation caused by predictive warning systems – The case of congestion tail warnings. *Transportation research part F: traffic psychology and behaviour*, 26, 49–61.

Naujoks, F., Wiedemann, K., & Schömig, N. (2017). *The importance of interruption management for usefulness and acceptance of automated driving*. Paper presented at the Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.

Niels, F., Edoardo, P., Florent, G. F., & Clément, V (2019). A new methodology to model driver behaviour accounting for the variation in driving manner using naturalistic driving data. In *Proceedings: 26th International Technical Conference on The Enhanced Safety of Vehicles.* Eindhoven, Netherlands.

OECD. (1990). *Behavioural adaptations to changes in the road transport system.* Organisation for economic co-operation and development, Paris.

Özkan, T.,& Lajunen, T. (2005). Multidimensional Traffic Locus of Control Scale (T-LOC): factor structure and relationship to risky driving. *Personality and individual differences*, 38(3), 533-545.

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

Pilutti, T., & Galip Ulsoy, A. (1999). Identification of driver state for lane-keeping tasks. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans.*, 29(5), 486–502. https://doi.org/10.1109/3468.784175

Portouli, E., Papakostopoulos, V., Lai, F. C., Chorlton, K., Hjälmdahl, M., & Wiklund, M. (2006). *D1.2.4: Long-term phase test and results.* Retrieved from http://www.aide-eu.org/pdf/sp1_deliv_new/aide_d1_2_4.pdf

Rudin-Brown, C. M., & Parker, H. A. (2004). Behavioural adaptation to adaptive cruise control (ACC): implications for preventive strategies. *Transportation research part F: traffic psychology and behaviour*, 7(2), 59–76.

Rudin-Brown, C., & Jamson, S. (2013). Behavioural adaptation and road safety: Theory, evidence and action. Boca Raton, FL: Taylor and Francis Group.

Saad, F. (2006). Some critical issues when studying behavioural adaptations to new driver support systems. *Cognition, Technology & Work,* 8(3), 175–181.

Saad, F., Hjälmdahl, M., Cañas, J., Alonso, M., Garayo, P., Macchi, L., Nathan, F., Ojeda, L., Papakostopoulos, V., Panou, M. & Bekiaris, E. (2004). *Literature review of behavioural effects (AIDE deliverable D1. 2.1).* Information Society Technologies, Brussels.

SAE (2021). *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems* (J3016): Society of Automotive Engineers.

Schmidt, G. J., & Rittger, L. (2017, September). Guiding driver visual attention with LEDs. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 279-286).

Skottke, E. M., Debus, G., Wang, L., & Huestegge, L. (2014). Carry-over effects of highly automated convoy driving on subsequent manual driving performance. *Human factors*, 56(7), 1272-1283.

Trösterer, S., Wuchse, M., Döttlinger, C., Meschtscherjakov, A., & Tscheligi, M. (2015, September). Light my way: Visualizing shared gaze in the car. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 196-203).

Van der Horst, A. R. A. (2013). Behavioural Adaptation to Roadway Engineering. In C. Rudin-Brown & S. Jamson (Eds.), Behavioural Adaptation and Road Safety. Boca Raton, FL: Taylor & Francis Group.

Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, 5(1), 1-10.

van Huysduynen, H. H., Terken, J., Meschtscherjakov, A., Eggen, B., & Tscheligi, M. (2017, September). Ambient light and its influence on driving experience. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 293-301).

Varotto, S. F., Farah, H., Bogenberger, K., van Arem, B., & Hoogendoorn, S. P. (2020). Adaptations in driver behaviour characteristics during control transitions from full-range Adaptive Cruise Control to manual driving: an on-road study. *Transportmetrica A: transport science*, 16(3), 776-806.

Victor, T. W., Tivesten, E., Gustavsson, P., Johansson, J., Sangberg, F., & Ljung Aust, M. (2018). Automation Expectation Mismatch: Incorrect Prediction Despite Eyes on Threat and Hands on Wheel. *Human Factors*, 60(8), 1095–1116. https://doi.org/10.1177/0018720818788164

Vogel, K. (2002). What characterises a “free vehicle” in an urban area? *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(1), 15-29.

Vogelpohl, T., Kühn, M., Hummel, T., & Vollrath, M. (2018). Asleep at the automated wheel—Sleepiness and fatigue during highly automated driving. *Accident Analysis & Prevention*, in press.

Wiedemann, K., Naujoks, F., Wörle, J., Kenntner-Mabiala, R., Kaussner, Y., & Neukum, A. (2018). Effect of different alcohol levels on take-over performance in conditionally automated driving. *Accident Analysis & Prevention*, 115, 89–97.

Wilde, G. J. (2013). Homeostasis Drives Behavioural Adaptation. In C. Rudin-Brown & S. Jamson (Eds.), Behavioural Adaptation and Road Safety. Boca Raton, FL: Taylor & Francis Group.

Wörle, J., Metz, B., Ottersen, I. & Baumann, M. (2020). Sleep in highly automated driving: Takeover performance after waking up. *Accident Analysis & Prevention 144:105617*. https://doi.org/10.1016/j.aap.2020.105617

Wörle, J., Metz, B., Thiele, C. & Weller, G. (2019). Detecting sleep in drivers during highly automated driving: the potential of physiological parameters. *IET Intelligent Transport Systems 13(8)*: 1241 – 1248.

Zhang, B., de Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation research part F: traffic psychology and behaviour*, 64, 285-307.

Zhang, T., Tao, D., Qu, X., Zhang, X., Lin, R., & Zhang, W. (2019). The roles of initial trust and perceived risk in public’s acceptance of automated vehicles. Transportation Research Part C: Emerging Technologies, 98, 207-220.

Annex

Mental model item list	Correct answer
1. The system works on motorways.	Yes
2. The system works on rural roads if lane markings are clear.	No
3. The system obeys priority rules.	No
4. The system drives up to 200 km/h in sections without speed limit.	No
5. The system works in construction sites.	No
6. The system overtakes slower lead vehicles.	Yes
7. The system gives a warning when the driver has to take back control.	Yes
8. The system brakes when the front vehicle brakes.	Yes
9. The sysem only works with clear lane markings.	Yes
10. The system works when the road surface is snowy.	No
11. If the driver does not respond to a take-over request, the system brakes to standstill in the current lane.	Yes
12. The system works in traffic jam and in slow-moving traffic.	Yes
13. The system works in villages and urban areas.	No
14. The system obeys the obligation to drive on the right.	No
15. The system works in all weather conditions.	No
16. I am allowed to sleep while the system is active.	No
17. I have to observe the surrounding traffic while the system is active.	No
18. I am allowed to close my eyes when the system is active.	Yes
19. I am allowed to do phone calls with the phone in my hands while the system is active.	Yes
20. I have to take back the vehicle control within a short time when the system asks me to do so.	Yes