

# Master Thesis

Evaluating two automation feedback aids' influence on the acceptance of automated truck merging systems

Michael de Jonge





# Master Thesis

Evaluating two automation feedback aids' influence on the acceptance of automated truck merging systems

by

Michael de Jonge

to obtain the degree of Master of Science  
at the Delft University of Technology,  
to be defended publicly on Tuesday March 12, 2019 at 14:00.

Student number: 4087224  
Project duration: February 5, 2018 – March 12, 2019  
Thesis committee: MSc. F. A. Dreger, TU Delft, daily supervisor  
Dr. ir. R. Happee, TU Delft, supervisor  
Dr. ir. J. C. F. de Winter, TU Delft  
MSc. J. C. J. Stapel, TU Delft

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.



# Abstract

*Introduction:* 35.000 people were killed in vehicle crashes in the US alone in 2016 [5], of which about 90% were to be attributed to human error [65]. Despite thorough regulations and driver training, large trucks are responsible for approximately 11.5% of all fatalities in traffic [6]. Merging is one of the most demanding tasks for a truck driver due to the size, weight and limited visual clearance of a truck. Automating the merging procedure can be the solution for a considerable number of accidents by removing the human factor. People have accepted the risk they are in when driving or being driven but acceptance of automated driving systems remains limited.

*Objective:* The goal of this study is to evaluate two automation feedback aids' influence on the acceptance of automated truck merging systems. This study looks at two types of feedback aids; the automation confidence level and a top view display where surrounding vehicles are visualized for the driver. The feedback aids are evaluated on the number of abortions, mental workload, trust in the system, perceived usefulness of the system and system satisfaction.

*Background:* Acceptance is the degree to which an individual incorporates the system in his/her driving, or, if the system is not available, intends to use it. Drivers of large trucks exhibit heightened criticism of automation due to professional identity, exposure or familiarity with the traditional task, and some degree of technical knowledge of the current system [17]. Introducing automation feedback gives the driver more information on the automation systems' status. 'The degree of certainty that the automation is able to handle the current situation' is the second most essential information in automated driving [8]. The presentation of the system's confidence level calibrates trust more appropriately [35]. The third and fourth most relevant information are the presence of surrounding vehicles and distances to them. An abstract visualisation of this information is shown on the top view display.

*Methods:* 41 participants experienced 12 simulations of a truck merging on the highway in a fixed base truck simulator. A high (1.5s gap) and medium workload (2.1s gap) condition paired with merging between cars or trucks created four simulation conditions to test each feedback aid condition (no feedback aid, confidence bar and top view). Mental workload was measured by tonic activity of the GSR. Trust, perceived usefulness and system satisfaction were measured with Jian's trust and Van der Laan's system acceptance questionnaires after each feedback aid condition [42, 73].

*Results:* Five participants rated their vulnerability as 'none at all', their measurements were removed from further analyses of the feedback aids. A total of 41 abortions were obtained from 432 simulations. A significant difference in abortion rate was found between no feedback aid (22 abortions), the confidence bar (13 abortions) and the top view display (6 abortions). A vulnerability threshold was seen between medium workload simulations (5 abortions) and high workload simulations (36 abortions). No abortion rate difference was found between merging between cars and trucks caused by visual obstruction. Participants' mental workload did not differ between feedback aid conditions, workload conditions or platoon conditions. Trust also showed no difference between the feedback aid conditions, this can be explained by the limited experience the participants gained with each feedback aid condition. Participants rated both feedback aids as more useful and satisfying than driving without a feedback aid with the top view display being favoured over the confidence bar. No correlations were found between driver's license years and trust, perceived usefulness or system satisfaction in any of the feedback aid conditions.

*Conclusion:* Both feedback aids increased the acceptance of automated truck merging systems. The number of abortions decreased and willingness to use increased by adding either the confidence bar or the top view display. The top view display showed the largest improvement in acceptance but more extensive research is needed in both feedback aids to assess their benefits in automated driving systems.



# Preface

*Before you lies my master thesis, it is the result of studying 7 years at the Delft University of Technology, 13 months of which were for my literature study on measuring trust in automation and research in the world of automated truck driving. It has been written to fulfil the graduation requirements of the Vehicle Engineering track as part of the Mechanical Engineering master at the 3mE faculty. The research was undertaken as contribution to the truck merging project at the Cognitive Robotics department.*

*Firstly, I would like to thank Felix Dreger for his extensive guidance and support during this process. The numerous meetings, comprehensive feedback on my literature study, my experimental design and interpretation of the measurements as well as the quick answers on my questions have been a tremendous help in the process of graduating. During my masters I have discovered an interest for the interaction between automation and user. Felix's knowledge of psychology and human-machine interaction has encouraged me to look critically at the design of automation and the way people use it.*

*Secondly, I would like to thank dr.ir. Riender Happee for his overview on the project, his opinions on my research and the discussions together with Felix and me. His help was indispensable in deciding the subject of my literature study and formation of the experimental design.*

*Thirdly, I would like to thank all the friends and fellow students who participated in the experiment. The time they took to help me with the experiment, the useful feedback they gave on the automation feedback aids and the overall enthusiasm for my work.*

*Lastly, I would like to thank my parents and my friends. They have supported and motivated me wherever possible during the highs and lows of my graduation process. They helped me to become the person I am now and I am really thankful for having them in my life.*

*Michael de Jonge  
Delft, February 2019*





# Contents

<b>Abstract</b>	<b>iii</b>
<b>Preface</b>	<b>v</b>
<b>List of Figures</b>	<b>ix</b>
<b>List of Tables</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Background</b>	<b>3</b>
2.1 Acceptance of automated driving systems . . . . .	3
2.1.1 Defining acceptance . . . . .	4
2.1.2 Automation Acceptance Model . . . . .	4
2.1.3 Measuring acceptance . . . . .	4
2.1.4 Trust . . . . .	5
2.2 State of the art truck automation. . . . .	6
2.3 Feedback systems. . . . .	7
2.3.1 Automation confidence feedback . . . . .	8
2.3.2 Top view display . . . . .	9
2.4 Merging procedure . . . . .	9
2.4.1 Defining the merging procedure . . . . .	9
2.4.2 Steering path. . . . .	10
2.4.3 Gap size . . . . .	11
<b>3 Methods</b>	<b>13</b>
3.1 Participants . . . . .	13
3.2 Experimental setup . . . . .	13
3.3 Simulations . . . . .	14
3.3.1 Road layout . . . . .	14
3.3.2 Truck. . . . .	14
3.3.3 Traffic . . . . .	15
3.4 Interface design. . . . .	16
3.4.1 On/off icon . . . . .	16
3.4.2 Confidence bar. . . . .	16
3.4.3 Top view display . . . . .	17
3.5 Procedure. . . . .	17
3.6 Measurements . . . . .	17
3.6.1 Demographic data . . . . .	17
3.6.2 Acceptance measurements . . . . .	18
3.6.3 Mental workload measurements. . . . .	18
3.7 Variables . . . . .	20
3.7.1 Independent variables . . . . .	20
3.7.2 Dependent variables. . . . .	20
<b>4 Results</b>	<b>21</b>
4.1 Prior ADAS experience . . . . .	21
4.2 Vulnerability . . . . .	21
4.3 Abortions . . . . .	22
4.3.1 Abortions split by feedback aid condition . . . . .	22
4.3.2 Abortions split by workload condition . . . . .	23
4.3.3 Abortions split by platoon condition. . . . .	23
4.3.4 Elapsed time at abortions . . . . .	24

4.4	Galvanic Skin Response data . . . . .	25
4.5	Questionnaire data . . . . .	27
4.5.1	Trust results . . . . .	27
4.5.2	Perceived usefulness & system satisfaction results . . . . .	28
<b>5</b>	<b>Discussion</b>	<b>29</b>
5.1	Acceptance . . . . .	29
5.1.1	Abortions . . . . .	29
5.1.2	Mental workload. . . . .	30
5.1.3	Trust . . . . .	30
5.1.4	Perceived usefulness & system satisfaction. . . . .	30
5.2	Feedback aid evaluation . . . . .	30
5.3	Limitations . . . . .	31
5.4	Further research . . . . .	31
<b>6</b>	<b>Conclusion</b>	<b>33</b>
	<b>Bibliography</b>	<b>35</b>
<b>A</b>	<b>Participant instruction form</b>	<b>41</b>
<b>B</b>	<b>Simulator screenshots</b>	<b>43</b>
<b>C</b>	<b>Experiment questionnaires</b>	<b>47</b>
C.1	Demographic questions. . . . .	47
C.2	Van der Laan's acceptance questionnaire . . . . .	47
C.3	Jian's trust questionnaire . . . . .	48

# List of Figures

2.1	The Automation Acceptance Model . . . . .	4
2.2	Aspects of defining appropriate trust in automation . . . . .	6
2.3	Four options for system confidence level information . . . . .	8
2.4	BMW's Surround View Assistant . . . . .	9
2.5	Merging procedure traffic overview . . . . .	10
2.6	Merging procedure segments . . . . .	10
3.1	Experimental setup . . . . .	14
3.2	Road layout in the simulation . . . . .	14
3.3	Simulator monitor visuals . . . . .	15
3.4	Simulator monitor visuals with added feedback aids . . . . .	15
3.5	Automation on/off icon . . . . .	16
3.6	Confidence bar design . . . . .	16
3.7	Top view display design . . . . .	17
3.8	TMSi Mobita and Ambu BlueSensor N ECG electrodes . . . . .	19
4.1	Adaptive Cruise Control and Automatic Lane Keeping use while driving . . . . .	21
4.2	Self-reported perceived vulnerability . . . . .	22
4.3	Simulation abortions split by feedback aid condition . . . . .	22
4.4	Simulation abortions split by workload condition . . . . .	23
4.5	Simulation abortions split by platoon condition . . . . .	23
4.6	Elapsed time at abortion . . . . .	24
4.7	Elapsed time at abortion split by workload and platoon conditions . . . . .	25
4.8	Galvanic Skin Response during 4 simulations of 1 participant . . . . .	25
4.9	Mean skin conductance for each feedback aid . . . . .	26
4.10	Mean skin conductance for each feedback aid in both workload and platoon conditions . . . . .	26
4.11	Jian's trust questionnaire response for each feedback aid . . . . .	27
4.12	Trust response by experience group . . . . .	28
A.1	Participants instruction form - page 1 . . . . .	41
A.2	Participants instruction form - page 2 . . . . .	42
B.1	Simulator instructions - part 1 . . . . .	43
B.2	Simulator instructions - part 2 . . . . .	44
B.3	Simulator instructions - part 3 . . . . .	45
C.1	Demographic questions before the start of the experiment . . . . .	47
C.2	Van der Laan's system acceptance questionnaire . . . . .	47
C.3	Jian's trust questionnaire . . . . .	48



# List of Tables

- 2.1 The Unified Theory of Acceptance and Use of Technology questionnaire . . . . . 5
- 3.1 Gap times and distances for both workload conditions . . . . . 15
- 4.1 Timing of simulation events . . . . . 24
- 4.2 Mean trust for each feedback aid . . . . . 27
- 4.3 Mean perceived usefulness for each feedback aid . . . . . 28
- 4.4 Mean system satisfaction for each feedback aid . . . . . 28





# Introduction

Every day billions of people transport themselves and others by car, bus, train and many other vehicles. People have accepted the risk they will be in when driving or being driving in these vehicles despite all the accidents that happen. In the US alone 35.000 people were killed and 2.2 million were injured in police-reported vehicle crashes in 2016 [5]. About 90% of which were to be attributed to human failure, for instance, through fatigue, inattention or drowsiness at the wheel [64, 65]. Large trucks are responsible for 9.25% of total vehicle miles driven in the US in 2015 where strict regulations and driver training result in only 3.67% large truck involvement in all accidents. However, fatal accidents have a large truck involvement rate of 11.48%, with 84% of those fatalities not being occupants of the truck [6]. This high fatality rate is caused by the large weight difference between trucks and other road users [23, 25]. Drivers of large trucks are a relatively neglected human factors subject in comparison to car users [60, 76, 80]. The commercial driver naturally spends more time behind the wheel in comparison to private vehicle drivers (up to 56 h in any given work week) [54], and the sustained mental workload associated with long-term tasks may cause performance to deteriorate [49]. Commercial vehicle drivers may exhibit, relative to private vehicle drivers, stronger stress reactions to traffic conditions and commit more risky driving behaviours [56], a factor which may compound itself in the time pressure which exists the industry.

Merging on the highway is one of the hardest parts of truck driving caused by the size and weight of the truck, the speed difference with other vehicles, slow acceleration and deceleration and large blind spots to the driver. Automating the merging procedure can be the solution for a considerable number of accidents by removing the human factor. Besides the safety aspect, automated truck merging systems can also make more efficient use of the road decreasing time spent at merging areas by 15% for all vehicles [67]. Clear communication between automated driving systems and the driver is very important to get the benefits from the joined system [61]. Besides clear operational instructions, the intention and user requirements from the system are needed to split the tasks between automation and operator. A mishap in this communication is seen in a Tesla accident in 2016 [4]. The car drove with 119 kilometres per hour into the side of a tractor trailer at an intersection. The car had the autopilot function turned on which keeps the car at the correct speed and position in the lane while the driver is required to remain vigilant and aware of the surroundings. The report stated the driver was not applying the brakes or started any manoeuvre to evade the trailer despite being able to see it for 7 seconds. The investigation report gives no decisive verdict but most likely the driver was not paying attention to the road and the system as is required and asked for by the autopilot function. This example shows a mismatch between the reliance on the system and the reliability of the system. The driver trusted the system so much he relied on it without performing his monitoring task.

The goal of this study is to evaluate two feedback aids' influence on the acceptance of automated truck merging systems. Increasing the amount of information given to the driver increases the mental workload but it can increase trust in the system, perceived usefulness and system satisfaction of using the system. The study looks at two types of added feedback information:

- Automation confidence level
- Top view image

The automation confidence level shows the confidence of the automated merging system in a successful and safe continuation of the merging manoeuvre. The top view image shows a visualisation of the own vehicle with the surrounding traffic and road layout as if a camera is located several meters above the vehicle pointing downwards. The visualisation is constructed from camera's and radar sensors in the vehicle. Automation confidence feedback can increase the understanding of automation error to mitigate the loss of trust in automation during high workload situations. This lets the driver know when the automation can be fully trusted and when he should be wary. There is a threshold which can't be crossed in terms of automation confidence, below this threshold the confidence feedback only decreases the acceptance of the system. Mental workload is another region of interest, adding more feedback information increases the mental workload with the enlarged task of monitoring the system but an increase in trust reduces the mental workload.

The hypotheses established for this study are:

1. Adding automation confidence feedback to automated truck merging systems reduces the number of abortions.
2. Adding a top view display to automated truck merging systems reduces the number of abortions.
3. The effect of both systems will be smaller in less critical conditions, operationalized as a larger gap (2.1s instead of 1.5s).
4. Adding automation confidence feedback to automated truck merging systems increases drivers' willingness to use the system.
5. Adding a top view display to automated truck merging systems increases drivers' willingness to use the system.
6. Adding automation confidence feedback to automated truck merging systems reduces drivers' mental workload.
7. Adding a top view display to automated truck merging systems reduces drivers' mental workload.
8. Automated merging between trucks yields more abortions than automated merging between cars.



# 2

## Background

This chapter gives some background for the research presented in this thesis. Section 2.1 elaborates on acceptance in general, acceptance of current Advanced Driver Assistance Systems (ADAS) and trust in automated driving. The state of the art in automated truck merging systems is discussed in section 2.2 and section 2.3 focuses on the visual feedback systems which can be added to vehicles. Lastly, section 2.4 provides information on the merging procedure, steering path and gap sizes.

### **2.1. Acceptance of automated driving systems**

Many companies are developing ADAS and autonomous driving vehicles but a high acceptance of the systems is required for widespread use. A poorly designed system inside vehicles can adversely affect the benefits associated with the system [9]. One of the most important characteristics of ADAS is to provide information to drivers. ADAS devices that are limited to supplying information are most likely to meet a priori acceptance. Examples are route guidance systems, traffic jam warning systems and driver monitoring systems. 24-33% of people are willing to use Fully Automated Driving (FAD) vehicles with younger people having a higher tendency to use them indicating an increase in willingness with time [1, 45]. The high a priori willingness is a positive sign for the upcoming industry, however, a high willingness to try new automation is correlated to people with the image of perfect automation. This image is quickly distorted once the automated system makes a mistake and turns out to be imperfect leading to a reduction of trust and disuse of the system [20]. This reduction of trust can only be restored with time or training. Drivers of large trucks exhibit heightened criticism of automation due to professional identity, exposure or familiarity with the traditional task, and some degree of technical knowledge of the current system [17], a finding which can be further supported by similar human factors research in air traffic control [10]. Handling over control to a system is evaluated as a negative aspect of ADAS [38]. The European SAVE project, aimed to develop a system that takes over vehicle control in case of real emergency, conveyed an international questionnaire survey that indicated that drivers are reluctant to release vehicle control but are willing to accept it in emergency situations [9]. Risk is the main factor in differences between automated driving and automation in other processes. A mistake in other processes (e.g. robotic arm in a manufacturing plant) can result in a reset of the system and some economic cost but a mistake in automated driving can result in a crash with relatively high economic cost or even fatalities. A difference with other high-risk automation (e.g. autopilot in an airplane, control of a nuclear reactor) is the limited reaction time in automated driving when automation fails and a lack of training with the automated systems. The high vulnerability paired with the short reaction time and lack of training requires a high level of trust from operators to accept and use the automated driving systems.

### 2.1.1. Defining acceptance

Acceptance is a scientific concept which is defined in many ways. Definitions of acceptance can be divided in five categories [2]:

1. Using the word "accept"
2. Satisfying needs and requirements
3. Sum of attitudes
4. Willingness to use
5. Actual use

Adell (2009) proposed a new definition built on these categories, focusing on a system's potential to realise its intended benefits, which is used to define acceptance in this report:

*Acceptance is the degree to which an individual incorporates the system in his/her driving, or, if the system is not available, intends to use it.*

### 2.1.2. Automation Acceptance Model

The Technology Acceptance Model (TAM) [13, 14] is a commonly used base model for modelling acceptance and together with compatibility and trust forms the Automation Acceptance Model (AAM) [32]. The AAM shows the process from External Variables to Actual Use of the System covering all five of the categories of acceptance definitions (figure 2.1). Reliance and acceptance are closely related as reliance is part of the fifth category of acceptance definitions; the actual use of the system.

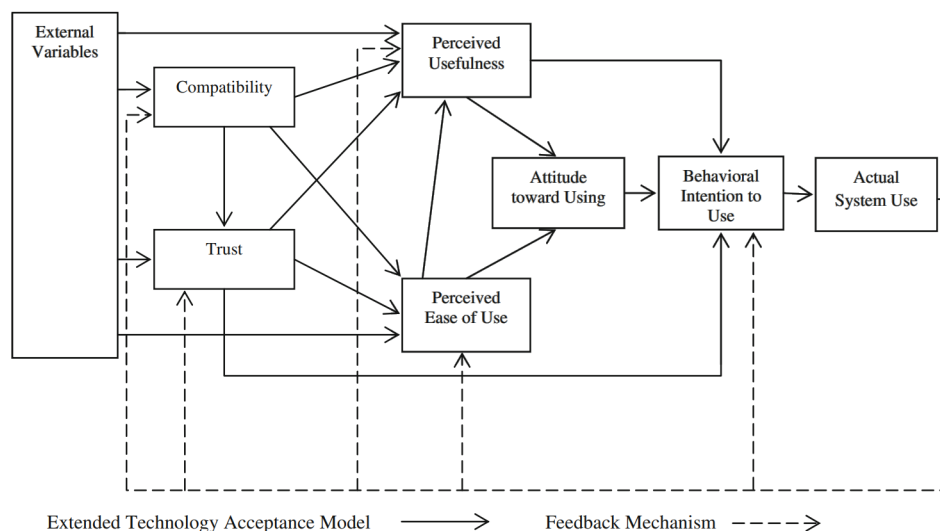


Figure 2.1: The Automation Acceptance Model [32]

### 2.1.3. Measuring acceptance

Measuring acceptance of ADAS or FAD cars is difficult. The great diversity in definitions of acceptance have created a similar diversity of measurements. Acceptance can be measured by focusing on one aspect of the AAM or one category of the five definitions of acceptance but more often it is measured by a combination of aspects of the AAM or categories of the definitions. A commonly used method of measuring acceptance is the Unified Theory of Acceptance and Use of Technology (UTAUT) questionnaire (table 2.1) [3]. It focuses on several aspects of the AAM resulting in a mean acceptance level. The UTAUT questionnaire is broadly used in acceptance research but is considered not specific enough to be used for acceptance of ADAS. The UTAUT questionnaire lacks questions about trust in the system, which is considered to be of great importance for FAD and ADAS acceptance research. In this study acceptance was measured by combining the use and willingness to use the automated merging system. The use of the system is measured by reliance on the system and willingness to use the system is measured by trust, perceived usefulness and system satisfaction.

Table 2.1: The Unified Theory of Acceptance and Use of Technology questionnaire

Behavioural intention to use the system (BI):	
BI1	I intend to use the system in the next months
BI2	I predict I would use the system in the next months
BI3	I plan to use the system in the next months
Performance Expectancy (PE):	
PE1	I would find the system useful in my job
PE2	Using the system enables me to accomplish tasks more quickly
PE3	Using the system increases my productivity
PE4	If I use the system, I will increase my chances of getting a raise
Effort Expectancy (EE):	
EE1	My interaction with the system would be clear and understandable
EE2	It would be easy for me to become skilful at using the system
EE3	I would find the system easy to use
EE4	Learning to operate the system is easy for me
Social Influence (SI):	
SI1	People who influence my behaviour would think that I should use the system
SI2	People who are important to me would think that I should use the system
SI3	The senior management of this business has been helpful in the use of the system
SI4	In general, the organization has supported the use of the system

#### 2.1.4. Trust

Trust is an important factor in the acceptance and use of automation therefore it is an important factor to look at in the development of automated driving systems. Trust in ADAS should be matched to the trustworthiness of the systems. Using FAD systems means reallocating the dynamic driving task to the system. Trust that the expected end result is achieved is a big factor in relying on automation transferring the responsibility from the human driver to the automated driving system. Many researchers see trust as the attitude towards using automation, this attitude creates an intention to use the automation leading to the behaviour of relying on automated driving systems [48]. Therefore reliance can be seen as the final part in the acceptance of automation. The intention to use the system and further the behaviour of relying on it are influenced by trust but also by perceived usefulness and perceived ease of use as well as compatibility between the automated system and the situation.

Trust is largely based on observation of the behaviour of the automation therefore prior to exposure, automation must already be trusted to a level where it will be used for trust to grow [51]. Relying on automation provides operators with an opportunity to observe how the automation works and thus, develop trust in the system. It is not possible for operators to observe automation unless they are relying on it, this may cause trust in the system fail to recover because it is no longer used. This voids the possible recovery of trust by an increase in reliability of the automation through developments of the system [46, 47]. To conclude; trust can be an unstable system, trusting a system leads to more trust and a lack of trust negatively influences the recovery when reliability has been improved [46]. This unstable system can be stabilized by measuring the trust level and adjusting the interaction between automation and user accordingly. A correct level of reliance can be achieved by influencing the system's reliability. A difference in the reliance on the systems and the reliability of the system can lead to misuse and disuse of automation [57]. The appropriateness of trust in a system is described by three factors; calibration, resolution and specificity [48].

*Calibration* is the correspondence between the humans trust in the system and the automation's capabilities. Trusting a system too much is called overtrust and can lead to misuse of the automation (e.g. using it when it should not be used, failing to monitor it effectively) while trusting a system too little and rely on self-confidence is called distrust and can lead to disuse of the automation (e.g. ignoring or turning off automation when it should be used). In figure 2.2 the diagonal dotted line represents good calibration while overtrust is above the line and distrust under it.

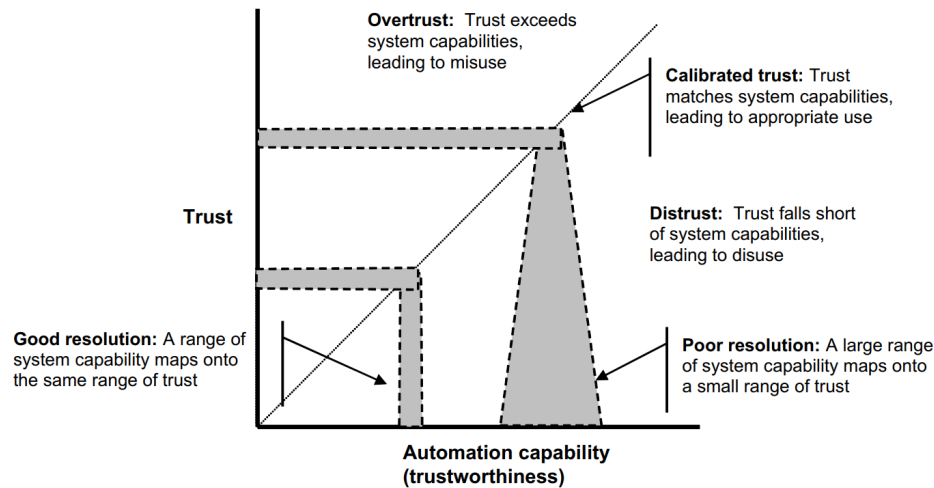


Figure 2.2: Aspects of defining appropriate trust in automation [48]

*Resolution* refers to the spread of the trust or capabilities, a systems reliability can vary over the extend of different situations whereas the user has a less dynamic trust level in the system therefore over- and distrusting the system in different situations.

*Specificity* is used for how specific the trust in the system is. Whether the user has a high trust in a small procedure of the automation or in the system as a whole. This last situation can be harmful when automation is working without errors 95% of the systems tasks but always fails at the other 5%.

The appropriateness of trust can be improved by giving feedback about the reliability of the automation. This feedback is often not given, leaving the operator with its own mental model of how the system works and what it is capable of even when it is not [21]. Adding performance feedback may resolve this problem but providing only the system's performance can mislead the user, lacking a comparison to their own performance. Adding feedback of the operator's performance can be very hard and is often unreasonable to do, especially in continuous non-dichotomous tasks. Training facilitates in the process of reducing inappropriate use of automation. Training reduces initial biases, provides knowledge about the systems capabilities and applies a risk assessment based on the behaviour of the automation [29]. Mercedes-Benz designed their Drive Pilot to slightly drift within the lane, not only to increase road comfort, but as a reminder that it is an assist system, not fully autonomous. "It's to make you stay engaged and aware of what's going on around you. It's letting you know that the car is assisting you, but it's not doing the entire job for you." according to Mercedes-Benz [33].

## 2.2. Sate of the art truck automation

The technical possibilities for a self-driving truck have been demonstrated in the Mercedes-Benz Future Truck 2025 project, but adjustments to the legal framework will be required to permit automated driving on public roads. The Vienna Convention on Road Traffic (1968) [52], which has been signed and ratified by almost all European States, is an international treaty designed to increase road safety by establishing standard traffic rules among the contracting parties. It states that the driver must at all times and in all circumstances be in control of his vehicle. UN/ECE Regulation R 79 for steering systems, which is based on the Vienna Convention on Road Traffic, permits corrective steering functions but does not permit automated steering at speeds in excess of 10 km/h. This is a restriction for HAD and FAD including an automatic merging system. Truck automation is restricted by the convention to ADAS like Adaptive cruise control with anti-collision systems, Lane keeping support, Lane changing support and driver alert support. Current FAD research for trucks including automatic merging focuses on highway platooning where multiple trucks drive in a convoy. This convoy uses Vehicle-to-Vehicle communication to drive at a very close distance having significant benefits for fuel consumption, road safety and the use of the existing road infrastructure [24].

The highway is the easiest place to start with FAD as there is a smaller variation in road users and a smaller range of permitted movements. The largest part of truck driving also takes place on highways at 70% of all kilometres (10% in residential areas and 20% on other roads) [72]. Despite the advantages of platooning drivers encounter a new driving situation due to the blocked field of view and the semi-automated driving at very close distance. New Human Machine Interfaces (HMIs) are being developed to overcome the issues drivers have with platooning (e.g. not able to react in time, not able to anticipate the evolving traffic situation). A study by Friedrichs suggests the use of continuous information visualization as a better traceability and hazard detection performance can be achieved [30]. It forms the basis for a good balance between the amount of information given and added mental effort for the driver as distracting a driver by too much information is a common mistake when designing HMIs. Developments in truck automation are lagging behind car automation caused by several factors. First, the car market is much larger, 14.6 million cars were registered in 2016 in Europe against 0.4 million trucks in the same year [55]. Second, trucks are commercial vehicles, the additional cost for ADAS and FAD systems can only be justified when they reduce other expenses or yield more revenue (e.g. more working hours allowed for drivers, multiple trucks per driver in platooning). Lastly, the size, weight and design of trucks complicates the development of hardware required for ADAS and automated driving systems (e.g. Bosch Automotive Steering introduced the first electro-hydraulic steering system worldwide for heavy commercial vehicles only in 2013 which enabled the implementation of ADAS in trucks [31]). Truck ADAS are focused on three aspects; keeping the correct distance to other vehicles (e.g. ACC and Emergency Brake), keeping within the lane (e.g. Lane Departure Warning) and driver monitoring (e.g. Driver Support Assistant). A merging operation requires systems to monitor distances to both the vehicle-in-front and vehicle on the left of the truck. Adaptive cruise control with emergency braking systems are available on most trucks to ensure safe spacing to the vehicle-in-front. Lane changing assistants, where a driver is warned when possibly colliding when changing lanes, are limited for trucks. No systems are available on the market to warn drivers or make corrective adjustments of the steering for vehicles on the ongoing lane. The only side vehicle warning systems available are Volvo's Lane Changing Support [71] and Mercedes-Benz's Sideguard Assist [70]. Both only warn when vehicles are in the blind spot on the right side of the truck. Development of fully autonomous vehicles is split in two different paths, traditional vehicle manufacturers (e.g. Tesla and Mercedes-Benz) develop, through ADAS, vehicles that take over an increasing number of tasks from the driver while other companies (e.g. Waymo and Uber) develop vehicles that start as fully autonomous vehicles without the need for a driver. Both parties do not have a commercially available vehicle able to merge automatically on the highway. Tesla released an 'On-ramp to Off-ramp' feature with version 9 of the Autopilot which guides a car from on-ramp to off-ramp with driver supervision but only when the car has merged on the highway [69]. Waymo has released a commercial version of their autonomous car, however, this only works in suburbs of Phoenix, Arizona excluding highway merging from the driving actions [27].

The level of automation is determined by the SAE levels, SAE level 0 is determined as the full-time performance by the human driver of all aspects of the dynamic driving task, even when enhanced by warning or intervention systems and SAE level 6 is determined as full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver [66]. The focus of this study will be on SAE level 2 automation where the merging procedure will be executed by the automation but the driver will be responsible for monitoring the driving environment and intervene if necessary.

## 2.3. Feedback systems

Introducing automation feedback gives the driver more information on the automation systems' state. However, care should be taken on designing the added automation feedback as monitoring it adds a monitoring task for the driver. Automation feedback is therefore a compromise between giving enough information for acceptance of the system and keeping the increase in mental workload by to the monitoring task to a minimum. This study focuses on two types of visual feedback, one feedback aid is showing the confidence the system has in a successful continuation of the automated driving and the other is a top view display showing the surrounding vehicles viewed from above the truck based on the truck's sensor data. A common rule in human interface design is that human machine interfaces should be an abstract visualisation of the data it represents. With the introduction of glass cockpits in airplanes the round altitude dial was changed to a vertical tape dial; a change in height was represented by a vertical tape moving up and down. Another example from the aviation industry is the pitch angle which is represented by a line moving up and down similar to the horizon going up and down when changing pitch angle.

After system status (i.e. active/inactive) and own speed, presence of surrounding vehicles and distance to them is the most relevant information to be given during highly automated driving when entering the highway [8]. Interview data from Beggiato showed that users expect to need less information over time due to higher trust. However, some information should never be removed which represents a hierarchy of essential information during automated driving from a user's perspective:

1. Status of the system (i.e. active/inactive, remaining time until take-over-request)
2. Degree of certainty that the automation is able to handle the current situation
3. Trip-related information: distance already driven and remaining distance/time
4. Current and planned manoeuvres
5. Current speed and current speed limits
6. Oncoming special/critical situations, e.g. construction zone, congestion

### 2.3.1. Automation confidence feedback

The algorithms of automated systems are getting more complex by time. Systems used to have only one input and one output and the system was either working fully correct or not at all. As the complexity of automation increases more tasks, humans are generally better in, are transferred to automation (e.g. deciding if a road user is a pedestrian or a cyclist). This creates uncertainty in the outcome of the automation understood as a system malfunction, significantly decreasing trust in the system. We have accepted other humans to make occasional mistakes but refuse automation to do so because we can't foresee when a mistake is more likely to occur and generally don't get feedback on why a mistake or unexpected action was made. A proven way to enhance the forgiveness on automated driving systems is by adding anthropomorphous to the car (e.g. give it a name and a voice) [28] but doesn't give insight of when or why the system failed. Although automation algorithms never doubt like humans do, the conformity of data from different sensors in combination with driving data, such as driving speed and information of the surrounding traffic, give the system data to calculate the chance of successful continuation. This feedback gives the driver more knowledge on when automation is more likely to be safe to use and when they need to be wary. This results in errors being corrected in time and more likely to be forgiven by the driver. Proof for the desire to know how confident the system is in successful continuation is found in Beggiato's study on essential information during automated driving from a user's perspective.

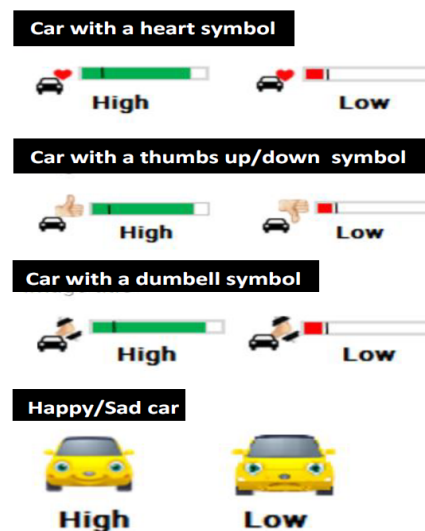


Figure 2.3: Four options for system confidence level information tested by Miglani (2017) [50].

Second to system status the most essential information is 'the degree of certainty that the automation is able to handle the current situation'. The implementation of confidence feedback tries to resolve two problems. First it reminds users that automation is not perfect and minor errors are as much part of automation as they are of human driving. Secondly it gives insight on what situations are difficult for the automation, when users should be wary and when they can relax. This helps conditional trust (i.e. know under what conditions the system can be trusted) to grow instead of unconditional distrust caused by minor errors in difficult situations. Unclear are the threshold values at which people no longer use the system because the confidence

levels shown are too low. Experiencing the system is needed for trust to restore, when a lack of trust prohibits the use of the system this will not happen. The data displayed is the chance of a successful action, this is not necessary a successful merge because this is not always possible. This can be abstracted to a confidence level of the automation. Just like self-confidence in human interaction, the system displays a self-confidence for successful progress of the automated driving. Confidence feedback is not commonly used, therefore no standard is available for giving feedback on the amount of confidence. Confidence also lacks a real-world representation (i.e. there is no way to see the confidence level of a person or automated system). A study performed by Miglani (2017) [50] tested four options for system confidence level information given to participants in an automated driving study (figure 2.3). More than half of the participants (53%) preferred to have a car with a thumbs up/down symbol and a bar level to represent the confidence level.

### 2.3.2. Top view display

Beggiato's research showed presence of and distance to surrounding vehicles as third and fourth most essential information to show during automated entering of the highway. A common rule in HMI design is that interface should be an abstract visualisation of what it represents. Drivers don't need to know exact distances to other vehicles or what specific vehicle it is. It is important to know if a gap is large enough to prevent a crash and it can be helpful to know if a vehicle is a car or a truck. A top view display gives the best visual representation of the surroundings as if a camera is located tenths of meters above the vehicle. The image is constructed using multiple cameras on the vehicle looking around. Algorithms transform the camera images from side to top view and stitch them together to form a surround view of the vehicle, an image of the top of vehicle is inserted to fill the top view display [62]. Similar to LIDAR technology the cameras can't see behind an object, especially when the object is as high or higher than the camera's position. Current systems show only a few meters outside of the vehicle to prevent misinformation by blocked camera views. The area visible on the displays is too small to help merging on the highway. BMW offers a Surround View display in their cars to help with parking. Figure 2.4 shows the image displayed to the driver made with a camera in the front, two on the mirrors and one on the rear of the vehicle. Developing these systems for trucks raises new issues, one is the position of the rear camera. Trucks change trailer often and a standard in trailers with cameras is needed for a surround display in trucks. Proceedings in development of this system have led to a system with surround view of the front and side of the cabin for pedestrian detection in low speed situations [22].



Figure 2.4: BMW's Surround View Assistant [39]

## 2.4. Merging procedure

### 2.4.1. Defining the merging procedure

The merging procedure involves careful adjustment of speed and coordinates of the vehicle to change lanes onto the highway. The differences in speed and gap distances between the merging and surrounding vehicles determine if a merging procedure can be executed safely and comfortably. An overview of the vehicles surrounding the merging vehicle can be seen in figure 2.5. To merge on the highway an automated vehicle is required to perform two partially decoupled tasks; 1) to adjust its speed so as to maintain headway safety and 2) to steer so as to control the lateral motions of the vehicle. Speed adjustment usually takes place on the

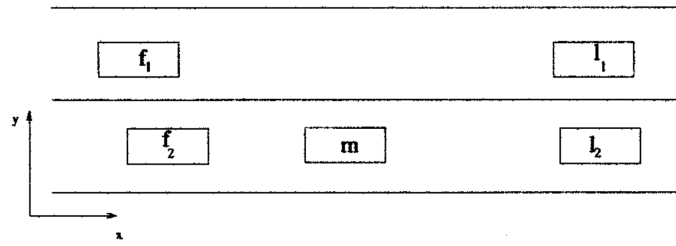


Figure 2.5: Merging procedure traffic overview

entering ramp and first part of the acceleration lane (e.g. the first 100 meters of the acceleration lane to have accelerated to at least 75% of the ongoing lane speed in the Netherlands [75]) while steering to change lanes starts halfway on the acceleration lane. Trucks are permitted to drive 80 km/hr on Dutch highways, however, because they have slow acceleration and most highways are designed for higher speeds acceleration to full speed or nearly full speed is recommended for safe merging. Guidelines for truck acceleration show a necessary 980 meters required to reach full speed [75] while 553 meters is the average road needed to reach full speed in heavy trucks [78]. Differences in speed between ongoing traffic and trucks are not uncommon, however, to keep the complexity of this study low, the acceleration lane is made long enough to accelerate to full speed and match the speed of the ongoing traffic. After the acceleration the lane changing manoeuvre takes place when a suitable gap has been found. The lane changing manoeuvre can be split in three parts; Head is the part in the lane of origin, Cross is the part where the truck is on the line and Tail is the part in the lane of destination. A graphical representation can be seen in figure 2.6. The mean times for heavy trucks are respectively 1.21s, 3.18s and 1.45s for head, cross and tail. Giving a mean total time of 5.84 seconds for the lane change [81]. With a speed of 80 km/hr the distance needed for the lane change procedure is approximately 130 meter.

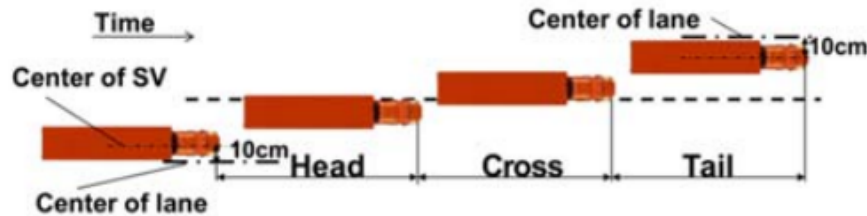


Figure 2.6: Merging procedure segments

### 2.4.2. Steering path

Successful implementation of automated lane changing systems requires a higher level of comfort and consistency to increase the acceptance. The forgiveness we have in ourselves and other drivers is mainly created by having feedback, discussing or knowing why a manoeuvre was jerky, very close to other traffic or abandoned halfway. To compensate the lack of interaction between driver and automation improving the automation's performance has always been a critical point (e.g. if people never feel unsafe they never have to doubt if a system is trustworthy). Lateral acceleration and jerk (i.e. change of lateral acceleration in time) are important parameters for the comfort experienced by the driver in a lane changing manoeuvre. Comfort is increased when both these parameters are decreased, so a steering path should be chosen to minimize these. The best way with minimum jerk and lateral acceleration is the trapezoidal acceleration trajectory where steering is started with maximum jerk, then at maximum lateral acceleration the jerk is kept to zero and all is reversed to steer back to parallel to the ongoing lane [12]. Jerk is kept to a minimum for transition times of 5,5s and longer. Transition times longer than 6,5s show little difference in jerk for the extra time added. Human drivers have a natural way of using the trapezoidal acceleration trajectory with a mean transition time of 5.84 s in natural driving [81].



### 2.4.3. Gap size

Gap size is an important aspect in automated merging because it creates vulnerability. Trust is needed for the acceptance of automation in vulnerable situations. When vulnerability is removed trust is no longer relevant to decision making since there is no or very little consequences to making the wrong decision [19]. A small gap is considered less comfortable and less accepted than a large gap to vehicles-in-front [15]. Larger gap sizes decrease the chance of hitting another vehicle when automation malfunctions so drivers will accept automation more easily. To merge, the gap, between the vehicles on the ongoing lane, should be large enough to fit the merging vehicle as well as a gap to the vehicle-in-front and a gap to the rear vehicle. Both post-merge gaps should be large enough for the speed difference between the vehicles and an emergency braking procedure. Other influences for deciding if a gap size is accepted or rejected include; weather, vehicle size-type, emotional status and number of leading vehicles on the ramp [44]. A differentiation is made between cooperative merging and forced merging. Cooperative merging involves the cars in the ongoing lane to change lane, speed up or slow down to accommodate for the merging vehicle. This study focuses on forced merging to decrease a feeling of shared responsibility with other drivers on the road [77].

*"The most desirable type of gap merge would be both 'optional' and 'ideal' in that it would not be made because the ramp driver had run out of acceleration lane nor would it cause turbulence in the freeway stream."* [18]

Performing the ideal merge requires a gap size that fits the merging vehicle and gaps to the front and rear vehicle after the merging procedure. Average headway times were measured in manually driven traffic. In high traffic flow car-to-truck headway averaged around 2.61s and truck-to-car headway averaged around 2.55s [79]. An ideal merge remains these headway times at the end of the merging procedure, this requires a total headway at the start of the merging procedure of 6.01s including 0.85s for the length of the merging truck. Normal headway times differ on the type of vehicle and the amount of traffic on the road. In a situation of uncongested very low traffic headway times of around 8.38 to 10.32 seconds can be seen between vehicles. With very high traffic the headway times lower to 1.95 to 2.75 seconds between vehicles [79]. With a speed of 80 km/hr this leaves a gap of 43 to 61 meters between the vehicles, large enough to merge into with slim safety margins. The minimum safety spacing for lane changing (MSSLC) is smallest when the speed difference is 0 or the vehicle moves away longitudinally from the merging vehicle. The smallest MSSLC with the vehicle-in-front is 35 meters and 22 meters to the rear vehicle [43]. Results for the MSSLC vary and depend among other things on the comfortable acceleration, time duration of the lane change and difference in speed between the vehicles. Adding the gaps to the truck length requires a total gap of 75,75 meters needed for a safe lane change. Gaps in high traffic are only 43 to 61 meters so a totally safe merging procedure is not possible without cooperative merging, however, this is often still performed. The Time-to-Collision (TTC) is the time it would take before the rear vehicle would make contact to the rear of the truck. Merges with the TTC smaller than 2 sec occur in 8.09% of Mandatory Lane Changes (MLC) and merges with the TTC smaller than 3 sec in 16.3% of MLCs [81]. Merging procedures with a small TTC are executed often in manual driving but are less self-evident to be acceptable by automated truck merging systems. Do we really permit automation to act in a hazardous way? This is a controversial question without a correct answer but it is likely necessary in order to accept automated systems and replace the human driver.



# 3

## Methods

This chapter gives an outline of the research methods used in the study. It provides information on participant inclusion criteria, the experimental design and design of the simulations and interface. Lastly, the procedure, measurements used for data collection and variables in the experiment are discussed.

The words 'simulation', 'section' and 'experiment' are used throughout the following chapters. A simulation is one video of a feedback aid condition in one scenario (e.g. a video with an added confidence bar in a high workload trucks platoon condition). A section consists of all four simulation for one feedback aid and the questionnaires at the end of the four simulations. The experiment is described as all three sections combined with the inclusion of the instructions at the beginning (i.e. the moment the participant enters the room until he/she leaves the room).

### 3.1. Participants

A total 41 persons participated in the study (27 male, 14 female) with little or no experience with automated driving systems. Mean age was 27.5 years (SD = 8.56), ranging from 21 to 59 years. All participants possessed a car driver's license for on average 9.3 years (SD = 8.41) ranging from 4 to 41 years and had normal or corrected-to-normal eyesight.

### 3.2. Experimental setup

The TU Delft's truck simulator was used in the experiment. The simulator consists of 2 TV monitors (65" and 50") which display the front and front/right side windows of the truck. One computer runs the experiment and the other, a dSpace virtual machine, was turned on to increase the fidelity of the experiment. The experiment was created using PsychoPy2 Experiment Builder (v1.90.3) [58, 59]. The steering wheel is from a DAF-truck powered by a Force Feedback SENSO-Wheel SD-LC/T, the pedals are powered by Force Feedback SENSO-Pedal. Both the SENSO-Wheel and SENSO-Pedals were turned off because the participants don't perform any manual driving actions. Participants were asked to keep their hands at the steering wheel to increase their involvement. A Kitbon Mini 18 Keys numeric keypad was attached to the middle of the steering wheel. The participants used the numeric keypad to advance in the instructions and to abort simulations. A laptop was placed on a stool beside the participant on the side of their dominant hand for answering the questionnaires. In trust and acceptance research high simulator fidelity is important to make the participant feel penalized for making a wrong decision. The simulator used has medium fidelity, missing movement and full immersion. The lack of high fidelity was compensated by shortening the gap sizes, increasing the vulnerability experienced by the drivers. Figure 3.1 shows the complete set-up.



Figure 3.1: Experimental setup

### 3.3. Simulations

Twelve simulations were created for the participants, 4 conditions for all of the 3 feedback aid conditions. The perception of automated truck driving was created by recording manually driven merges and showing participants these videos. A simulation was recorded for each of the 4 conditions and feedback aids were edited in the videos afterwards. The simulation starts with the ego-truck on the access ramp at 0 km/hr. The truck accelerates to 80 km/hr on the access ramp and continue on the acceleration lane. On the acceleration lane the truck merges in 5.5 to 6 seconds to the ongoing lane on the left after which it follows the road.

#### 3.3.1. Road layout

The road layout consists of a four-lane highway in two directions (figure 3.2, number 1, 2 and 3). One of the directions has an on-ramp consisting of an access ramp (number 5) continuing in an acceleration lane (number 4). The road is modelled after Dutch road standards, the lane width is 3.5 meters with a 3 meter wide emergency lane on the right [75].

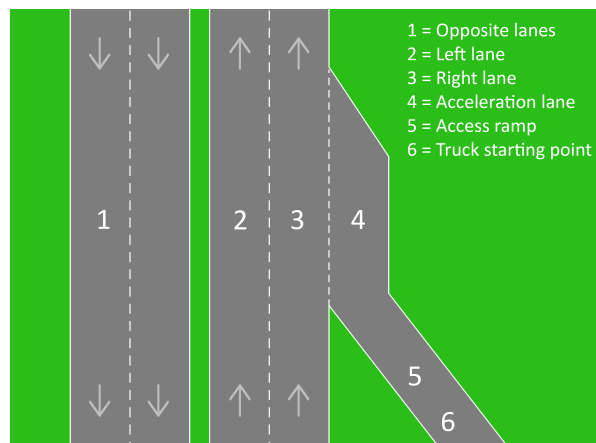


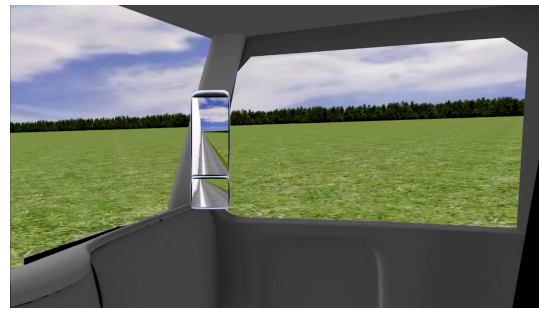
Figure 3.2: Road layout in the simulation

#### 3.3.2. Truck

The truck used is a standard 40t truck with a 13.6m box trailer, the total length of the vehicle is 18.75m which is the standard maximum allowance in the EU for these vehicles [53]. Longer trucks are allowed in some EU countries up to 25.25m but for this test the 18.75m variant is used. The width of the truck is 2.5m and it is dynamically modelled after a fully loaded 40 ton truck.



(a) Front monitor visuals



(b) Right monitor visuals

Figure 3.3: Simulator monitor visuals

The standard visuals for the front and right monitor can be seen in figures 3.3a and 3.3b. The front screen in figure 3.3a has no added feedback aids whereas figure 3.4a and figure 3.4b show the visuals with the confidence bar and top view display added respectively.



(a) Front monitor visuals with added confidence bar



(b) Front monitor visuals with added top view display

Figure 3.4: Simulator monitor visuals with added feedback aids

### 3.3.3. Traffic

The maximum allowed speed for trucks in the Netherlands is 80 km/hr [74]. On the right lane a platoon of ten vehicles was created with a constant gap size between them driving at 80 km/hr, the same speed as the ego-truck during the merging procedure, to merge between. The truck merged between the third and fourth vehicle in all scenarios. The left lane was kept empty as it interfered with the measurements needed to validate the desired gap sizes. The opposite lane had some cars on it but was very quiet in general.

#### Gap size conditions

Two workload conditions were created by simulating two different gap sizes. A medium workload situation, in normal traffic, has a headway of 3.73 seconds on average and a high workload situation a headway of 2.61 seconds on average [79]. Headway includes the length of the vehicle so this relates to gap times of 2.89 seconds and 1.77 seconds respectively. The low fidelity of the simulator was compensated by decreasing the gap times to increase mental demand on the participants. The 2.1 seconds gap will be referred to as 'medium workload condition' and the 1.5 seconds gap will be referred to as 'high workload condition' (table 3.1). The other vehicles are adaptive so after the merging procedure the vehicles behind the truck slow down to restore the set gap size.

Table 3.1: Gap times and distances for both workload conditions

Workload	Gap time	Gap distance
Medium	2.1 s	46.7 m
High	1.5 s	33.3 m

### Platoon conditions

Two different vehicle platoons were created, one with only cars and one with mixed traffic where vehicle three and four were always trucks. The cars platoon is referred to as 'cars platoon condition' while the mixed platoon is referred to as the 'trucks platoon condition'. Trucks create a larger visual obstruction and gap sizes tend to be larger between trucks (e.g. 1.75s headway between cars and 2.75s headway between trucks in high traffic flow [79]). It is expected that in otherwise similar conditions the trucks platoon condition yields more abortions. The two platoon conditions combined with the two workload conditions form four simulation combinations used to test each feedback aid condition.

## 3.4. Interface design

### 3.4.1. On/off icon

System status (active/inactive) is the most essential information for a driver when using automated driving systems [8]. The system status is presented by an on/off icon on the dashboard (figure 3.5). The automation is turned on in every simulation and never turned off. For lack of a standard icon for automated drive mode an icon has been designed for this experiment following general HMI design rules (e.g. green/red for on/off). The icon is placed in the middle of the instrument cluster between the tachometer and the speedometer.

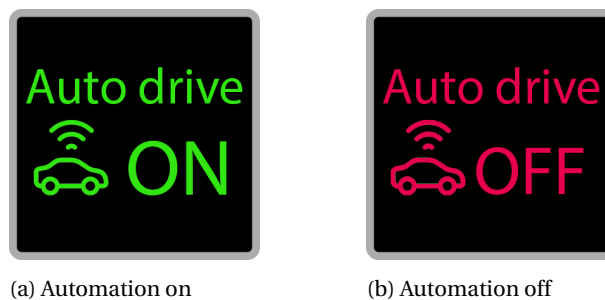


Figure 3.5: Automation on/off icon

### 3.4.2. Confidence bar

The confidence bar is modelled after Miglani's research in confidence feedback interfaces [50]. The high confidence condition has a thumb up and a green bar while the low confidence condition shows a thumb down and a red bar, in between the bar turns orange with a thumb up. The confidence bar remains static during the simulations at 95% full to be certain that simulations aren't aborted because the vulnerability threshold is crossed by the confidence feedback. Participants experience a moving confidence bar in the crash simulation video they see at the start of the confidence bar section.

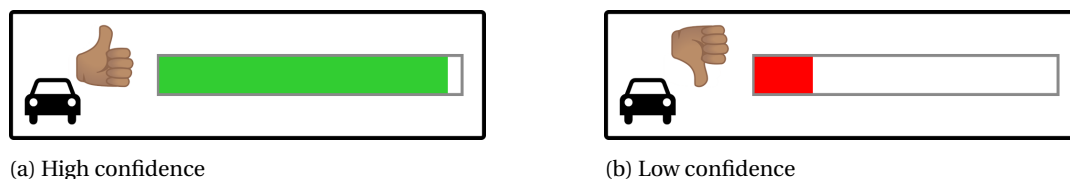


Figure 3.6: Confidence bar design

### 3.4.3. Top view display

The top view display designed for this study is the image from a virtual camera located about 50 meters above the truck. The image visible to the driver is about 70 by 70 meter with the ego-truck in the middle. This perfect top view display is not feasible with current technology, however, with further development of hardware and algorithms this top view display is a compromise between present-day top view displays and future technical improvements.

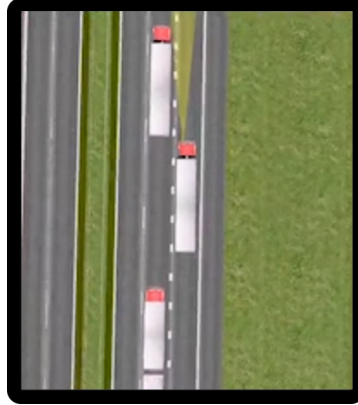


Figure 3.7: Top view display design

## 3.5. Procedure

After arriving, participants received written instructions about the experiment and signed the informed consent form. Two electrodes were placed on the inside of their non-dominant hand on the index and middle or ring finger. The resistance between the electrodes was measured and needed to be between  $200k\Omega$  and  $900k\Omega$ . Hereafter the participants took place in the simulator and were asked to fill in the first part of the questionnaire containing the demographic questions (appendix C.1). Once the participants were connected to the TMSi Mobita the simulator was started. Participants first received instructions on what they would see and what they would need to do. They also saw a sample simulation to get familiar with aborting the simulator. The sample simulation they saw is the medium workload cars scenario with no feedback aid. After the instructions the experimental simulations started. The participant received four simulations for one of the feedback aids (cars and trucks scenario both with medium and high workload gap times). After the four simulations the participants were asked to fill in both Jian's trust questionnaire and Van der Laan's system acceptance questionnaire on a laptop beside them. When the questionnaires were completed the simulations of the next feedback aid condition started, until all three were completed. To cancel out learning effects the feedback aids as well as the scenarios were simulated in a random alternating sequence. After the final questionnaires participants answered how vulnerable they felt during the experiment after which they were thanked for their participation and the electrodes were removed. The simulations and answering of the questionnaires took around 10 minutes for each feedback aid. Combined with the instructions and placement of the electrodes the experiment required around 45 minutes. The participant instruction form can be found in appendix A and simulator instructions can be found in appendix B (screens 1 to 14).

## 3.6. Measurements

### 3.6.1. Demographic data

Participants answered the following demographic questions at the start of the experiment:

- What is your age?
- What is your gender?
- How many years do you have a driver's license?
- How often do you use Adaptive Cruise Control while driving?
- How often do you use Automatic Lane Keeping while driving?

### 3.6.2. Acceptance measurements

The acceptance of the automated merging systems is measured by looking at the simulator abortions, a trust scale questionnaire and an acceptance scale questionnaire. The simulator abortions represent the use of the system and the willingness to use the system is represented by the trust, perceived usefulness and system satisfaction measurements.

#### Simulator abortions

Participants were instructed to abort the simulation when they considered the situation to be unsafe. The involvement was increased by multiple warnings that the simulator would crash when two vehicles touch each other and the experiment needed to be reset when this occurs. Each simulation can be aborted 10 seconds after the start until the end by pressing the '0' key on the numeric keypad. A press on the '0' key stops the simulation immediately and the next simulation is started. The PsychoPy experiment saves at which simulations the '0' key is pressed and at what time from the start of the video it was pressed.

#### Trust measurements

Measuring trust can be difficult, it has no general definition and only the consequences of trust, in combination with other influences, can be seen and measured. Questionnaires have long been the standard way of measuring trust, ranging from single trust questions "How much do you trust the system" to well thought out questionnaires like Jian's 12 item questionnaire [42]. Questionnaires provide the best representation of the complete aspect of trust where physiological and behavioural data mainly focus on the performance component of trust. A shortcoming in the information received from questionnaires is a lack of comprehensive feedback on what components have the biggest influence on the level of trust measured. Physiological and behavioural measurements are only recommended when questionnaires are not suitable for the experiment or as a backup to validate the correctness of the questionnaire. Eye tracking used to measure monitoring frequency has the best selectivity towards trust in automated driving in combination with little primary-task intrusion. Head movement and change in lighting condition should be kept to a minimum to increase the chance of correct measurements making it less ideal outside of simulator studies. When eye tracking is not possible, secondary task performance can be used to measure monitoring ratios to provide information on reliance. The other physiological measurements are not recommended because they focus on the fight-or-flight response of the body influenced by many factors whereof trust is only a small part. Questionnaires are suitable for this study as there is time to answer them between the sections. Jian's questionnaire is chosen because of its thorough research in developing it and its elaborate view on trust. The questionnaire can be found in appendix C.3.

#### Perceived usefulness and system satisfaction measurements

Not only should drivers feel safe and therefore trust the automated merging system users should also want to use a system. Assessment of the liking of a system can be done by one of many methods which has been developed for it. A questionnaire specialised in advanced transport telematics is the system acceptance scale developed by Van der Laan [73]. Nine simple 5-point rating-scale items form the questionnaire. Each item consists of 2 opposing properties between which the participant has to choose. These items load on two scales, a scale denoting the usefulness and a scale denoting the system satisfaction. Items 1, 3, 5, 7 and 9 are used for the usefulness scale and items 2, 4, 6 and 8 are used for the satisfactory scale. The questionnaire can be found in appendix C.2.

### 3.6.3. Mental workload measurements

Despite automation being developed to reduce mental workload, studies have shown an increase in the mental workload of the operator caused by the enlarged task of monitoring the system [7, 57, 61]. Mental workload measurements provide information about the increase in mental workload by the enlarged monitoring task as well as the trust participants have in the automated merging systems. A reduction in mental workload indicates high levels of trust while an increase in mental workload indicates low levels of trust. Self-reported mental workload data is considered most reliable [16] however mental workload should be measured during each simulation in this study.

A secondary task forces participants to split their effort between two tasks, it creates a measurement for how important the goal of one task is compared to the other task. In automated driving research the division of mental effort between monitoring the system and the secondary task can be interesting. An increase in trust results in a decrease in the need for monitoring therefore more effort is placed on fulfilling the secondary task. Secondary tasks are usually visually, aurally and/or cognitively demanding depending on how



much work is needed to force the driver to split his cognitive capacity between the tasks. Similar to reliance, simulator fidelity has to be high enough to get a realistic sense of vulnerability for correct measurements of secondary-task performance. Instructional design is a key aspect in secondary-task performance because the importance of one task over the other can easily be determined by the explanation of the experiment leaning toward automation monitoring or toward the secondary task completion. Secondary task performance is more a measurement of trust than mental workload as it keeps the mental workload constant and measures only the division between the primary and secondary task. As real-life use of the system does not have a secondary task and boredom is also a region of interest, a secondary task is not the best measurement of mental workload.

Physiological data is preferred since it can be measured continuously during the experiment. Heart rate is the most useful given that rest-measurements are used to assess baseline heart activity [16]. However, trials are too short for heart rate to give an accurate result, therefore the Galvanic Skin Response (GSR) is used in this study. GSR, or electrodermal activity, refers to the electrical changes in the skin as a result of activity of the Sympathetic Nervous System (SNS) [68]. Electrodes are placed on the skin a few centimetres apart and a small electric signal is used to measure the conductance of the skin between them. Sweat is created when sweat glands in the skin are stimulated by activity in the SNS, the ions in the secreted fluid increase the skin conductance [41]. This reaction is highest in the palm of the hand, fingers or the sole of the foot where the density of sweat glands is highest. GSR consist of two components; a tonic level of activity and the phasic activity [34]. Tonic activity, or Electrodermal Level (EDL), is the baseline activity, it gives an average GSR to the general state of the person. Phasic activity, or Electrodermal Response (EDR), shows the response on external stimuli. EDL is mainly used for measuring how a person is experiencing the scenario/experiment as a whole while EDR is used to measure psychological response to an experimental condition or a scenario event.

The GSR was measured using a Mobita physiologic signal amplifier system from TMSi. Two Ambu BlueSensor N ECG electrodes are placed on inside of the index and middle finger at the proximal phalanx bone. The resistance between the two electrodes should be between  $100k\Omega$  and  $1000k\Omega$ . The resistance was measured at the start of the experiment in a relaxed state, the experimental stimuli only decreases the skin conductance so a smaller bandwidth of  $200k\Omega$  to  $1000k\Omega$  was used. If the resistance was below  $200k\Omega$  gel was removed from the electrodes and/or the electrode on the middle finger was moved to the ring finger. Data from the Mobita was sent wireless to a laptop where it was stored with OpenVibe software. OpenVibe measured the signal at 1000Hz and stored it in a CSV-file. Figure 3.8a shows the Mobita and figure 3.8b the Ambu BlueSensor N ECG electrodes.



(a) TMSi Mobita



(b) Ambu BlueSensor N ECG electrodes

Figure 3.8

## **3.7. Variables**

### **3.7.1. Independent variables**

A within-subject study was performed varying 3 different variables; feedback aids, workload conditions and platoon conditions. Three different feedback aids combined with two workload conditions and two platoon conditions form a 3-by-2-by-2 study. Twelve simulations were presented to the participants formed by three feedback aid blocks, presented in random order, each containing the four simulation conditions in random order.

### **3.7.2. Dependent variables**

The dependent variables are; number of abortions and elapsed time at the abortions; galvanic skin response during each simulation, trust, perceived usefulness and system satisfaction for each feedback aid and lastly the perceived vulnerability of the experiment.

# 4

## Results

In this chapter the results of the study are presented. The ADAS experience, vulnerability, abortion, GSR, trust, perceived usefulness and system satisfaction data have been categorised by feedback aid conditions. The abortion and GSR data also have been split in both workload conditions and platoon conditions to check for a significant difference in specific scenarios.

### 4.1. Prior ADAS experience

Participants were asked how often they use Adaptive Cruise Control and Automatic Lane Keeping while driving. They answered on a 6-point scale ranging from 1 (Never) to 6 (Always). More than 65% of the participants never used Adaptive Cruise Control and 83% of the participants never used Automatic Lane Keeping (figures 4.1a and 4.1b).

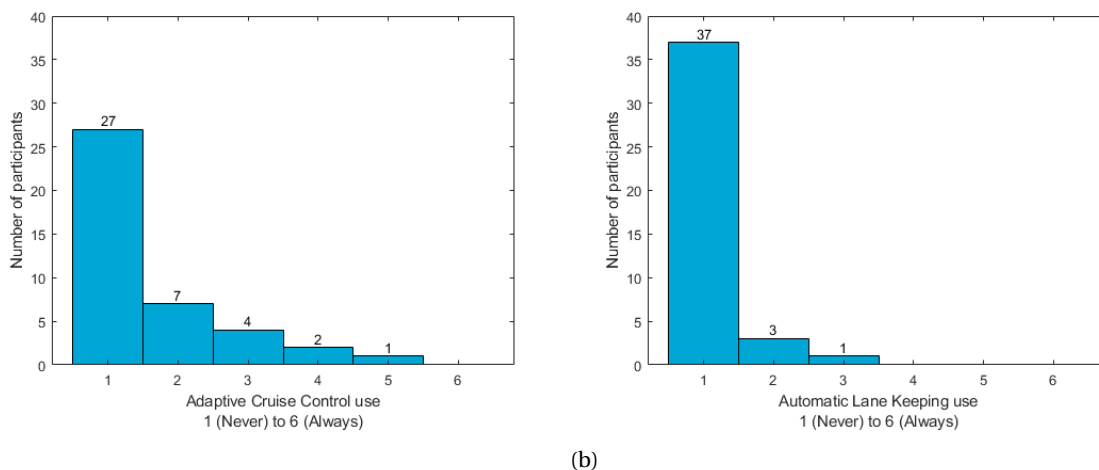


Figure 4.1: Adaptive Cruise Control and Automatic Lane Keeping use while driving

### 4.2. Vulnerability

Participants rated their perceived vulnerability at the end of the experiment on a 7-point scale ranging from 1 (Not at all) to 7 (Extremely). As discussed in section 3.2 vulnerability is needed for trust to matter in automation (i.e. if there is no penalty on wrong decision making there is no need for trust). Five participants rated their vulnerability to be 'none at all', their data was removed from further analyses of the feedback aids. The need for perceived risk is seen in the abortions rate of these participants, their abortion rate was 3.3%, much lower than the 9.5% average.

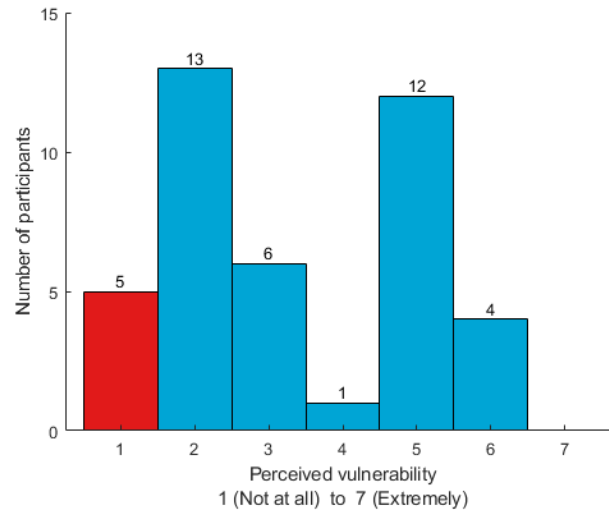


Figure 4.2: Self-reported perceived vulnerability (red indicates the removed participants)

### 4.3. Abortions

The participants were instructed to abort the simulation when they considered the situation to be dangerous or when they expected their vehicle to collide with another vehicle. Participants aborted the simulation 41 times in total (approximately 9.5% of all 432 simulations). A learning effect was expected, assuming participants quickly learned they experienced the same simulations in each feedback aid condition or that every simulation ends in a safe merge. The learning effect was very limited with the first, second and third section yielding 16, 12 and 13 abortions respectively.

#### 4.3.1. Abortions split by feedback aid condition

The abortions were categorised by feedback aid; no feedback aid, confidence bar and top view display (figure 4.3). No feedback aid had 22 abortions, the confidence bar had 13 abortions and the top view display had 6 abortions out of 144 simulations each. A binomial test indicated the ratio of abortions with the confidence bar at 0.090 to be lower than the ratio of abortions with no feedback aid at 0.153,  $p = .019$  (1-sided). Binomial tests also indicate the ratio of abortions with the top view display at 0.041 to be lower than both the ratio of abortions with no feedback aid at 0.153,  $p < .001$  (1-sided), and the ratio of abortions with the confidence bar at 0.090,  $p = .021$  (1-sided).

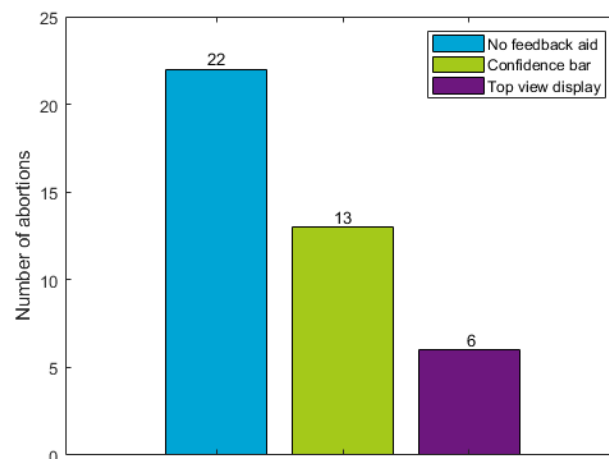
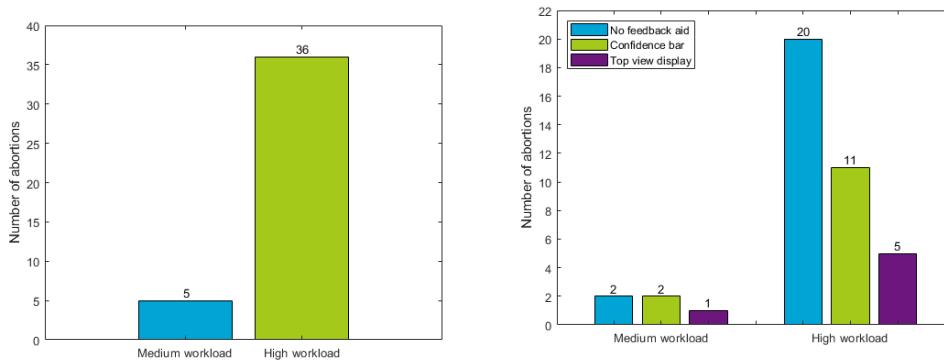


Figure 4.3: Simulation abortions split by feedback aid condition

### 4.3.2. Abortions split by workload condition

The abortions were split between the medium workload condition, with a 2.1s gap, and the high workload condition, with a 1.5s gap (figure 4.4a). The medium workload condition had 5 abortions and the high workload condition had 36, out of 216 simulations each. A binomial test indicated the ratio of high workload abortions at 0.167 to be higher than the ratio of medium workload abortions at 0.023,  $p < .001$  (1-sided). The abortions were further split by feedback aid for each workload condition (figure 4.4b). With 5 abortions out of 216 simulations there are insufficient data points to validate a difference between the feedback aid conditions in the medium workload condition. The high workload condition had sufficient abortions and binomial tests show the ratio of confidence bar abortions at 0.153 to be lower than the ratio of no feedback aid abortions at 0.278,  $p = .010$  (1-sided) and that the ratio of top view display abortions at 0.069 is lower both than the ratio of no feedback aid abortions at 0.278,  $p < .001$  (1-sided) and the ratio of confidence bar abortions at 0.153,  $p = .027$  (1-sided).



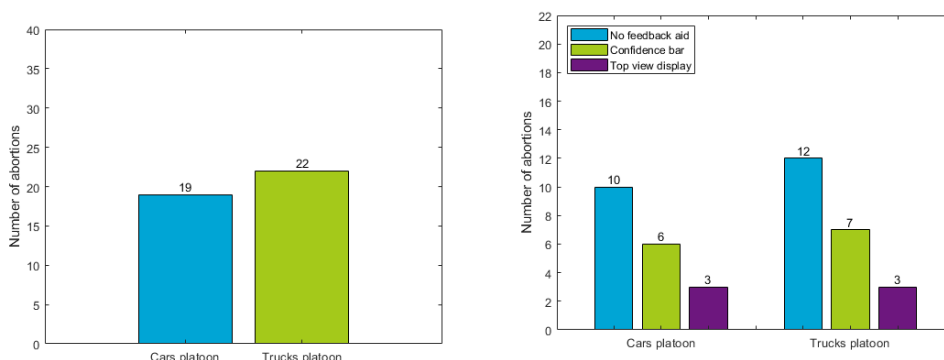
(a) Abortions split by workload condition

(b) Abortions split by workload condition and feedback aid condition

Figure 4.4: Simulation abortions split by workload condition

### 4.3.3. Abortions split by platoon condition

The abortions were split between the cars platoon condition, the truck merges between two cars, and the trucks platoon condition, the truck merges between two trucks (figure 4.5a). The cars platoon condition had 19 abortions while the trucks platoon condition had 22 abortions, out of 216 simulations each. A binomial test indicated the ratio of trucks platoon abortions at 0.102 not to be significantly higher than the ratio of cars platoon abortions at 0.088,  $p = .268$  (1-sided). The abortions were further split by feedback aid for each platoon condition (figure 4.5b). The distribution of abortions over the feedback aids is similar for both platoon conditions showing no indication for a different preference of feedback aid in one of the platoon conditions.



(a) Abortions split by platoon condition

(b) Abortions split by platoon condition and feedback aid condition

Figure 4.5: Simulation abortions split by platoon condition

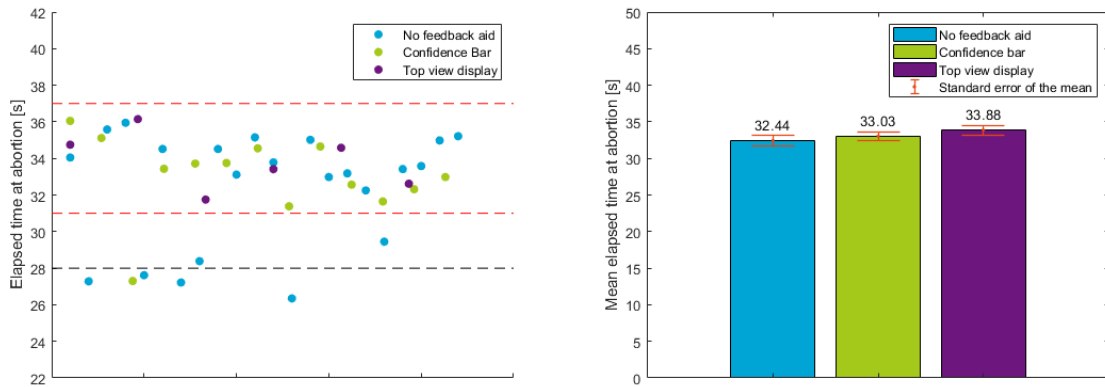
#### 4.3.4. Elapsed time at abortions

The simulation videos are 50 seconds of length, table 4.1 shows a list of the events during a simulation. The abortion data points are plotted by the elapsed time at abortion in figure 4.6a. The black dashed line marks the moment the truck enters the acceleration lane, driving side by side with the ongoing traffic, and the red dashed lines mark the start and end of the merging manoeuvre.

Table 4.1: Timing of simulation events

Time [s]	Event
0	Start acceleration on access ramp
15	First sight of other traffic
28	Entering acceleration lane
31	Start merging manoeuvre
37	End merging manoeuvre
50	End of simulation

The seven data points below the bottom dashed red line are abortions before the merging procedure started, many of which before driving on the acceleration lane (i.e. the simulations were aborted before the truck was driving side to side with the ongoing traffic). These early abortions have 4 trends; 1) they were all in the high workload condition, 2) they were all within the first 6 simulations experienced by the participants (median = 4th simulation), 3) they were all but one done by participants younger than the average participant (mean = 23.3 years) and 4) all but one was without feedback aids. A combination of young drivers, without feedback aids, in a high workload situation and with little experience with the automated merging system causes premature abortions.

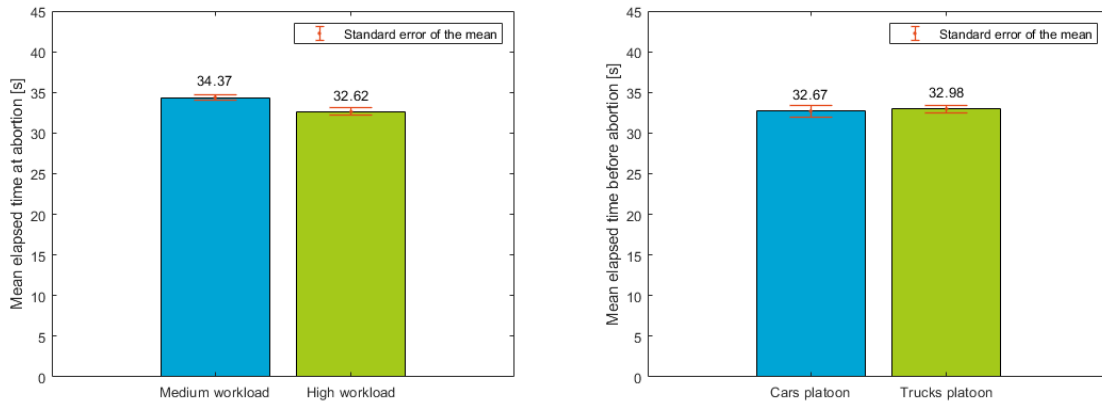


(a) Individual abortions by elapsed time at abortion

(b) Mean elapsed time at abortion for each feedback aid

Figure 4.6: Elapsed time at abortion

The mean elapsed time data for the feedback aids (figure 4.6b) can't be considered normally distributed according to Shapiro-Wilk's test as 'no feedback aid' is not normally distributed,  $p = .002$ . Friedman's Two-Way Analysis of Variance retains the null-hypothesis, stating no difference in elapsed time between the feedback aids,  $p = 1.000$ . The mean elapsed time data for the workload conditions (figure 4.7a) can't be considered normally distributed according to Shapiro-Wilk's test,  $p = .126$  and  $p = .001$ . The Independent-Samples Mann-Whitney U Test showed no significant difference in the median of elapsed times between medium workload and high workload,  $p = .139$ . The mean elapsed time data for the platoon conditions (figure 4.7b) can't be considered normally distributed according to Shapiro-Wilk's test,  $p = .001$  and  $p = .019$ . The Independent-Samples Mann-Whitney U Test showed no significant difference in the median of elapsed times between the cars platoon and the trucks platoon,  $p = .548$ .



(a) Elapsed time at abortion split by workload condition (b) Elapsed time at abortion split by platoon condition

Figure 4.7: Elapsed time at abortion split by workload and platoon conditions

### 4.4. Galvanic Skin Response data

The GSR was obtained by measuring the resistance between two electrodes on the inside of the index and middle or ring finger. The skin conductance data of 6 participants was removed as they were incorrect or incomplete due to malfunction of the TMSi Mobita or human error in starting the recording in time, leaving skin conductance data of 30 participants for analyses. A typical GSR measured over a period of four simulations is presented in figure 4.8. The green asterisks represent the start of a simulation and the red asterisks the end of a simulation. The magenta asterisks represent the start of a merging procedure and the blue asterisks represent the end of a merging procedure. The high workload scenario with cars surrounding the truck in this data set was prematurely aborted by the participant and emitted the highest skin conductance reaction.

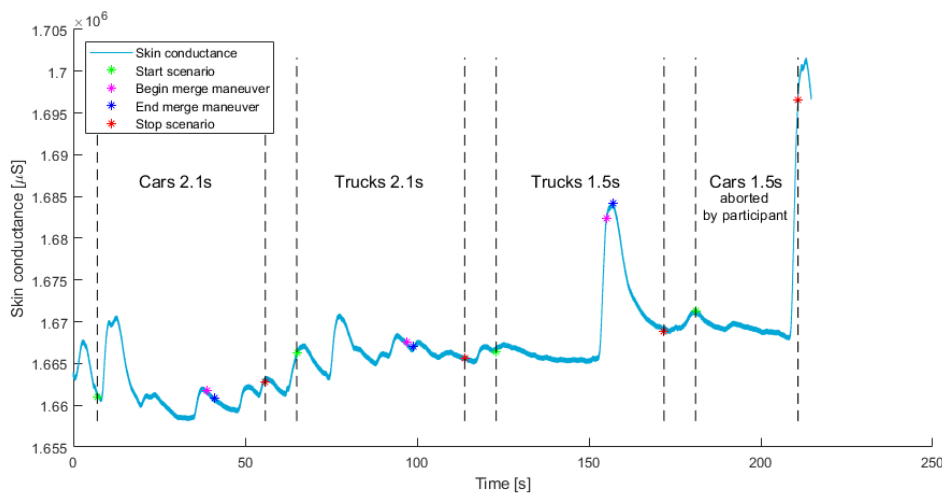


Figure 4.8: Galvanic Skin Response during 4 simulations of 1 participant

The Ledalab tool for Matlab was used to extract the tonic activity from the skin conductance data using the Continuous Decomposition Analysis [11]. Simulations aborted by the participant contain less data points, missing the data points after the merging procedure. These are mainly low mental workload data points which give a skewed representation when comparing mean skin conductance of aborted and non-aborted simulations. A mean skin conductance value for each scenario was created using the tonic activity data from 15 second after the start of a simulation, the point in time where traffic is seen for the first time, till 35 seconds after the start (slightly earlier than the end of the merging procedure to compensate for the aborted simulations). This reduces the difference in data between the mean skin conductance of aborted and non-aborted simulations.

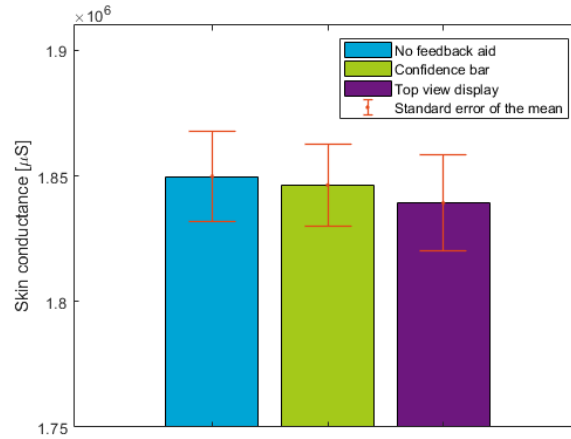
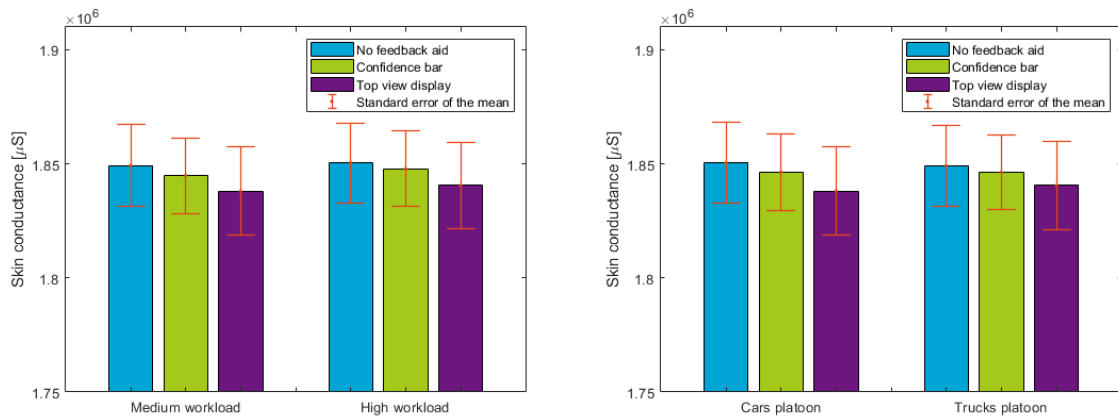


Figure 4.9: Mean skin conductance for each feedback aid

The skin conductance data for each feedback aid condition can be considered normally distributed according to Shapiro-Wilk's test,  $p = .056$ ,  $p = .276$  and  $p = .108$ . Mauchly's Test of Sphericity indicates that the assumption of sphericity has not been violated,  $\chi^2(2) = .746$ ,  $p = .689$ . A repeated measures ANOVA test shows no difference in skin conductance between the feedback aids to be statistically significant,  $F(2, 58) = 1.295$ ,  $p = .282$ . Figure 4.9 shows a plot of the mean skin conductance for each feedback aid and the standard error of the mean.



(a) Mean skin conductance in both workload conditions

(b) Mean skin conductance in both platoon conditions

Figure 4.10: Mean skin conductance for each feedback aid in both workload and platoon conditions

The GSR data was split by workload conditions and by platoon conditions to check for a significant difference between the feedback aid conditions (figures 4.10a and 4.10b). The medium workload data can't be considered normally distributed according to Shapiro-Wilk's test as the no feedback aid condition does not have a normal distribution,  $p = .039$ . Friedman's test retains the null hypothesis, which states that the distributions of GSR data for the feedback aids in medium workload conditions are the same,  $p = .195$ . The high workload feedback aids data can be considered normally distributed according to Shapiro-Wilk's test,  $p = .069$ ,  $p = .279$  and  $p = .092$  and did not violate the assumption of sphericity,  $\chi^2(2) = 1.894$ ,  $p = .388$ . No significant difference is present between the mean skin conductance of each feedback aid conditions in the high workload condition,  $F(2, 58) = 1.207$ ,  $p = .307$ . The cars platoon feedback aids data can be considered normally distributed according to Shapiro-Wilk's test,  $p = .062$ ,  $p = .211$  and  $p = .096$ , and did not violate the assumption of sphericity,  $\chi^2(2) = 1.618$ ,  $p = .445$ . No significant difference is present between the mean skin conductance of each feedback aid conditions in the car platoon condition,  $F(2, 58) = 1.733$ ,  $p = .186$ . The trucks platoon data can't be considered normally distributed, the 'no feedback aid' condition yields a signifi-



cance of  $p = .034$  in the Shapiro-Wilk's test. Friedman's test retains the null hypothesis, which states that the distributions of GSR data for the feedback aids in truck platoon conditions are the same,  $p = .067$ .

## 4.5. Questionnaire data

### 4.5.1. Trust results

Jian's trust questionnaire consists of 12 statements, 5 statements are negative (e.g. the system is deceptive) and 7 statements are positive (e.g. the system is reliable) [42]. The answers on the 5 negative statements were inverted so all answers range from 1 (Not at all) to 7 (Extremely). A mean trust level was calculated by averaging the 12 answers for each participant.

Table 4.2: Mean trust for each feedback aid

Feedback aids	Mean	SD	N
Trust with no feedback aid	4.34	0.78	36
Trust with confidence bar	4.44	0.97	36
Trust with top view display	4.55	1.09	36

The mean trust levels for each feedback aid were calculated by averaging the mean trust levels of each participant (table 4.2). A repeated measures ANOVA is used to test for a statistically significant difference between the mean trust levels of the feedback aids. The steps, from 1 to 7, in answering the trust questionnaire are considered to be equidistant, as only answers 1 and 7 were labelled, which is a requirement for the repeated measures ANOVA. Mauchly's test did not indicate any violation of sphericity therefore no correction is needed,  $\chi^2(2) = .841, p = .657$ . The difference between the means is not statistically significant:  $F(2,70) = .662, p = .519$ . Distribution of the trust results was checked but no significant differences were found (figure 4.11).

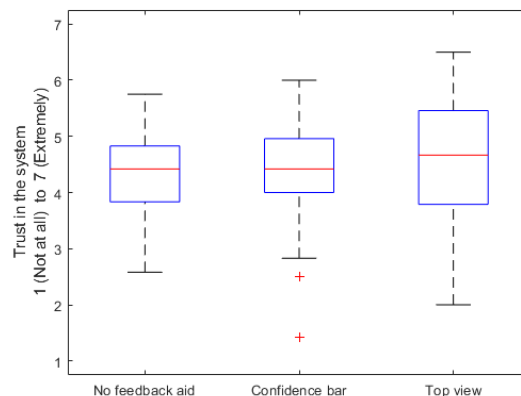


Figure 4.11: Jian's trust questionnaire response for each feedback aid

The amount of experience in driving manually can have an effect on the trust people have in the system and the amount and type of information favoured. The trust data was split in two equal groups of 18 participants, an inexperienced group with 4 to 6 years of driver's license years and an experienced group with 7 and more years of driver's license years (figures 4.12a and 4.12b). Experienced drivers tend to have a greater variance in their trust in the feedback aids indicating a greater division between users in the benefit of adding feedback aids.

A correlation between the trust and the number of abortions was investigated. Pearson correlation tests showed no correlation to exist between trust in the system and the number of abortions for each driver aid condition,  $p = .731, p = .378$  and  $p = .647$ .

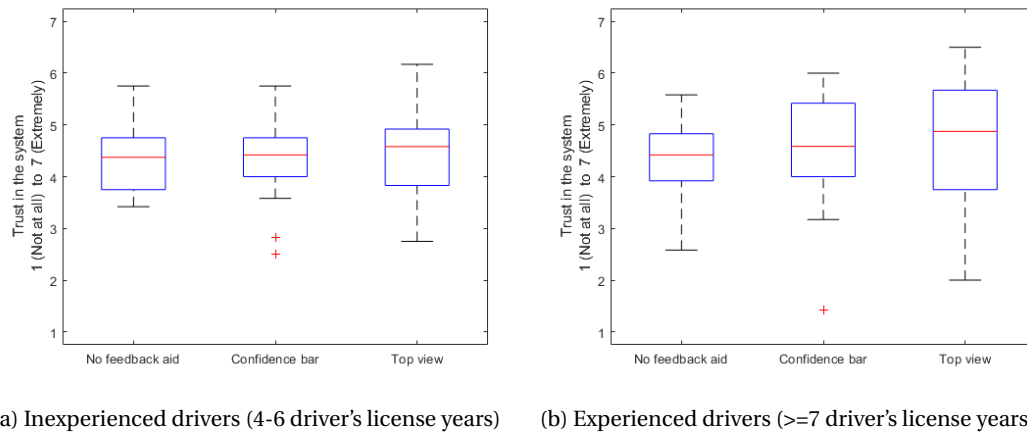


Figure 4.12: Trust response by experience group

#### 4.5.2. Perceived usefulness & system satisfaction results

Van der Laan's questionnaire consists of 9 statements [73]. The answers of questions 1, 2, 4, 5, 7 and 9 are inverted to make them consistent (1 = Not at all and 5 = Extremely). A mean perceived usefulness level was calculated by averaging the answers on statements 1, 3, 5, 7 and 9 and a mean system satisfaction level was calculated by averaging the answers on statements 2, 4, 6 and 8.

##### Perceived usefulness

The mean perceived usefulness for each feedback aid condition was calculated by averaging the perceived usefulness of all participants (table 4.3). A repeated measures ANOVA is used to validate if the difference in mean usefulness levels is significant. Mauchly's test of sphericity shows a violation of sphericity therefore a correction is needed,  $\chi^2(2) = 6.407, p = .041$ . A rule of thumb proposed by Howell and Field state that Huyn-Feldt results should be used when Greenhouse-Geisser epsilon is larger than 0.75 [26, 40]. The Greenhouse-Geisser epsilon is 0.853 therefore the Huyn-Feldt results are used. The difference between the mean usefulness levels is statistically significant:  $F(2, 70) = 4.379, p = .020$ . No significant correlations were found between driver's license years and perceived usefulness of the system in any of the feedback conditions,  $p = .339, p = .572$  and  $p = .339$ .

Table 4.3: Mean perceived usefulness for each feedback aid

Feedback aids	Mean	SD	N
Perceived usefulness with no feedback aid	3.39	0.69	36
Perceived usefulness with confidence bar	3.59	0.59	36
Perceived usefulness with top view display	3.78	0.65	36

##### System satisfaction

The mean system satisfaction for each feedback aid condition was calculated by averaging the system satisfaction of all participants (table 4.4). A repeated measures ANOVA is used to validate if the difference in mean satisfaction levels is significant. Mauchly's test of sphericity shows no violation of sphericity,  $\chi^2(2) = 5.167, p = .076$ . The difference between the mean satisfaction levels is statistically significant,  $F(2, 70) = 4.686, p = .012$ . No significant correlations were found between driver's license years and system satisfaction in any of the feedback conditions,  $p = .250, p = .932$  and  $p = .485$ .

Table 4.4: Mean system satisfaction for each feedback aid

Feedback aids	Mean	SD	N
System satisfaction with no feedback aid	3.42	0.80	36
System satisfaction with confidence bar	3.61	0.76	36
System satisfaction with top view display	3.90	0.79	36

# 5

## Discussion

The study's goal was to evaluate the influence of two automation feedback aids on the acceptance of automated truck merging systems. Two feedback aids have been developed and tested; a confidence bar showing the automation's confidence in a successful and safe continuation of the automated driving and a top view display showing a birds-eye view of the truck and surrounding traffic.

### 5.1. Acceptance

The acceptance was determined by four measurements; the number of simulation abortions, the level of trust, the level of perceived usefulness of the system and the level of satisfaction from the system. The trust level, perceived usefulness and system satisfaction together form the willingness to use the system.

#### 5.1.1. Abortions

Results show a significant decrease in the number of abortions by adding either of the feedback aids with the top view display showing the least number of abortions. No feedback aid, confidence bar and top view display having 22, 13 and 6 abortions respectively. Participants interpreted the instructions, on when to abort a simulation, in different ways. Some participants aborted the simulations when they had a minimal feeling of unsafety, resulting in 2 participants aborting over half of the simulations, while others only would abort when they were absolutely sure a crash would occur, resulting in 19 participants not aborting a simulation at all. One participant commented afterwards "I hoped the simulation would crash", despite instructing them not to let the truck crash at all cost, indicating the large difference in perceived vulnerability between simulator studies and real-life studies. A vulnerability threshold can be seen between the workload conditions, the medium workload has only 5 abortions compared to 36 abortions in the high workload condition. The three participants who aborted the medium workload simulations had double to triple the abortions in high workload simulations each as a result of the different vulnerability thresholds. The high workload abortions have a similar distribution over the feedback aid conditions as the total abortions. The distribution on the medium workload condition however has too little data points for a significant difference. Visual obstruction was also a region of interest, tested by merging between cars, with good visual clearance, and between trucks, where sight is limited by the size of the truck in front. The cars platoon condition had 19 abortions and the trucks platoon condition 22 abortions. The expectation that with no feedback aid a large difference between the cars and trucks condition would be seen, caused by visual obstruction [36], which would be nullified with the addition of feedback aids was not found in this study.

The elapsed time at which participants aborted the simulation was recorded and checked for differences between the feedback aid conditions. Higher acceptance rates were expected in later abortions as higher acceptance creates a higher threshold for allowed vulnerability. No significant difference was present in the elapsed abortion time between the feedback aid conditions. Seven simulations were aborted before the truck entered the acceleration lane, these abortions, as expected, were all made by younger than average participants in high workload conditions without a feedback aid and within the first 6 simulations. This shows a vulnerability threshold difference between sight of the gap at the end of access ramp or driving alongside the traffic on the acceleration lane.

### 5.1.2. Mental workload

The GSR data showed no differences in mental workload between the feedback aid conditions. Adding automation feedback normally increase mental workload with the enlarged monitoring task [7, 57, 61]. Some participants mentioned the feedback aids required too much attention therefore they liked the condition with no feedback aids the most. The lack of an increase in mental workload is a positive indication for the feedback aids as the decrease in mental workload caused by an increase in trust nullifies the added mental workload of monitoring the feedback aids. A general increase in skin conductance over the duration of the experiment was observed and some participants mentioned their hands getting sweaty from holding the steering wheel for prolonged periods of time. This trend increased the variance in the data contributing to a lack of significance.

### 5.1.3. Trust

Trust is seen as the attitude towards using automation, this attitude creates an intention to use the automation [48]. An increase in trust in the system was expected with the addition of feedback aids as they help participants to verify the system's actions. Trust can also be negatively influenced by the feedback aids, the feedback aids are generated by the system and can be misinterpreted or not be correct at all, leading to a decrease in trust. No significant difference was found in mean trust values between the feedback aid conditions. Trust in automation is a slow growing property, although trust can quickly decrease when automation fails [20], time and experience are needed to increase trust in automation. The limited amount of time and number of simulations spent with each system combined with the similarity between the systems explains the lack of difference in trust levels. Trust levels were compared between inexperienced drivers (4-6 driver's license years) and experienced drivers ( $\geq 7$  driver's license years). The distribution of trust response showed no significant differences between inexperienced and experienced drivers in mean trust levels. A few participants reported they did not trust the confidence bar because it was static which is very improbable. The presentation of automation confidence calibrates trust more appropriately [35]. This requires the automation confidence presented to be appropriate to the situation, which a static bar is unlikely to be. Other participants reported they did not trust the top view display as there were two red cars in front of the trust, on the top view display only one car was shown which appeared too far away to be the closest of the red cars. Therefore the 'missing' car was reported as a malfunction of the top view display.

### 5.1.4. Perceived usefulness & system satisfaction

Closely related to the definition of acceptance is the scale of system acceptance by Van der Laan. It is subdivided in a perceived usefulness and a system satisfaction scale. A significant difference was obtained between the perceived usefulness of all three feedback aid conditions with the top view display being favoured over the confidence bar (mean = 3.39, 3.59 and 3.78 for no aid, confidence bar and top view display). Some participants reported the confidence bar to be useless because it was static but it shows in general it is more useful to have it static than not having it. System satisfaction showed the same trend with no feedback aid giving the least satisfaction and the top view display giving the most (mean = 3.42, 3.61 and 3.90 for no aid, confidence bar and top view display). Many participants reported they liked the top view display so much they would like it in their manually driven car as well.

## 5.2. Feedback aid evaluation

The feedback aids were evaluated through data and through participants' reactions. The confidence bar, modelled after Miglani's design [50], was clear and positively perceived by the participants. The confidence bar was intended to indicate a trustworthy system. The confidence threshold was unknown as was the effect of variance in the confidence level, therefore the confidence bar remained static. It is very improbable that the system always has the same amount of confidence in the situation resulting in some participants not trusting the confidence bar or deeming it useless. The level of confidence can be calculated as a mean of the speed of the ego-vehicle, the surrounding vehicles and the conformity of the sensor data from the truck together with some possible other factors including the weather. The top view display was received with mixed attitudes, it was found to be the most useful as distance to other vehicles was given in a clear and easy way. Almost all participants mentioned after the experiment that they liked the top view display the most. A problem with a true top view display is the way distances are perceived as a driver. Due to the speed of the vehicle in longitudinal direction, distances in longitudinal direction are perceived much smaller than distances in lateral direction [37]. A top view display without correction therefore appears to show vehicles further away

in longitudinal direction than as they are perceived by the driver by looking out the window. Drivers with this knowledge can greatly benefit from this as the top view shows real distances but as we are used to rely on perceived distances as seen through the window it creates mistrust and confusion among inexperienced drivers.

### 5.3. Limitations

Experienced truck drivers were favoured as participants for this study as they are experienced with driving a large truck on the road and have a good understanding of the dynamics of a fully loaded truck. Recruiting truck drivers without financial compensation was not feasible therefore people with a car driver's license for 4 years or more were recruited. A minimum required group size of 30 participants was estimated for this in-between to obtain significant results. Despite the recruitment of 41 participants in total, some expected results were not obtained due to large variance or too little data points. The vulnerability threshold which determines if a participant aborts a simulation differed considerably between participants. Two participants had a low threshold and aborted 7 and 8 simulations while more than half of the participants did not abort a single simulation. Careful planning of the instructional design did not help to remove this variation.

Results show the necessity of high vulnerability situations for feedback aids to make a difference. The participants however have not been subjected to a crash simulation for various reasons. The first is a lack of creating a realistic crash situation, vehicles are able to drive inside each other therefore simulator fidelity is lost when an experienced crash is not realistic or has no real penalty. Second, the simulations for each feedback aid condition were exactly the same, only varying in order, to make the best comparison between the feedback aids. Participants are very good at noticing when they see the same simulation again, this would only be increased when an important event (e.g. a crash) occurs in one of the simulations. Lastly, crashes occur very rarely, only 1 crash with injuries occurs every 5 million kilometres travelled [6]. This study focuses on the first experience with feedback systems during automated truck merging, therefore safe driving conditions were chosen.

The final limitation is the amount of experience the participants gained with each system. An in-between design study was chosen to limit the number of participants required combined with the limited time requested from the participants lead to a limited amount of time with each system from which the measurements were obtained. Trust is a slow growing aspect so no differences between the feedback aid conditions for trust in the system were obtained.

### 5.4. Further research

This study is one of the first in automated truck merging looking at the automation feedback aids. A clear difference between the automated merging system without a feedback aid and the automated merging system with a feedback aid was obtained. Future research should lead to improvements of both feedback aids; creating an algorithm for a dynamic confidence bar and increasing the real-world feasibility as well as solving the perception distortion regarding longitudinal distance of the top view display. The effects of long-term use of any of the feedback aids is also a region of interest to be explored, especially the confidence bar as it has the potential to give drivers more insight on when the system can be trusted and when it can't, improving knowledge of the specificity in trustworthiness of the system. Operators tend to trust the automation if the basic working/reasoning can be understood and appear capable of achieving the operator's goals in the current situation [63].



# 6

## Conclusion

In conclusion, this study compared the acceptance of automated merging systems without feedback aids to automated merging systems with either an automation confidence bar or a top view display. Acceptance was stated by the definition of Adell (2009): *Acceptance is the degree to which an individual incorporates the system in his/her driving, or, if the system is not available, intends to use it.* The hypotheses, as proposed in the introduction, are discussed and verified or rejected where possible.

1. Adding automation confidence feedback to automated truck merging systems reduces the number of abortions.
2. Adding a top view display to automated truck merging systems reduces the number of abortions.

Without feedback aids the simulator was aborted 22 times. The addition of a confidence bar resulted in 13 abortions and 6 abortions were obtained with the addition of a top view display. Results show a significant decrease in abortions with the use of either automation confidence feedback or a top view display with the top view display showing the greatest reduction. Hypotheses 1 and 2 are validated by this study

3. The effect of both systems will be smaller in less critical conditions, operationalized as a larger gap (2.1s instead of 1.5s).

The larger, 2.1s, gap simulations yielded only 5 abortions, no significant difference between the 3 feedback aid conditions could be obtained from the limited abortion data points. The lack of abortions indicate there is less need for feedback aids in medium and low workload situations. Hypothesis 3 can't fully be confirmed due to a lack of data.

4. Adding automation confidence feedback to automated truck merging systems increases drivers' willingness to use the system.
5. Adding a top view display to automated truck merging systems increases drivers' willingness to use the system.

The willingness to use the system was defined by a combination of trust in the system, perceived usefulness and system satisfaction. Although trust did not differ between the feedback aids both the confidence bar and the top view display had an improvement in perceived usefulness and system satisfaction confirming hypotheses 5 and 6.

6. Adding automation confidence feedback to automated truck merging systems reduces drivers' mental workload.
7. Adding a top view display to automated truck merging systems reduces drivers' mental workload.

Mental workload data showed no differences between the feedback aid conditions. Although the expected decrease was not obtained, no increase in mental workload caused by the expanded task of monitoring was not present. Hypotheses 7 and 8 are not validated by this study

8. Automated merging between trucks yields more abortions than automated merging between cars.

Automated merging between trucks yielded 22 abortions and merging between cars yielded 19 abortions. The difference was not significant so both conditions have a comparable abortion rate. The distribution of abortions over the feedback aid conditions was similar for both workload conditions to the general distribution. Hypothesis 9 is rejected based on the results of this study.

To conclude, this study is a continuation on the research in the use of automation confidence feedback and top view displays as part of human machine interaction. It is the first time these systems are evaluated on their influence on the acceptance of automated truck merging systems. The decrease in abortions and increase in willingness to use indicate the possible benefits of these feedback aids in automated truck merging. The total amount of abortions in this study shows the need for more research in the interaction between autonomous vehicles and their users. More research is required to perfect human machine interaction before automated driving is accepted in our daily life, increasing safety on the road, decreasing workload for truck drivers and increasing the efficient usage of our congested roads.



# Bibliography

- [1] Hillary Abraham, Chaiwoo Lee, Samantha Brady, Craig Fitzgerald, Bruce Mehler, Bryan Reimer, and Joseph F Coughlin. Autonomous vehicles, trust, and driving alternatives: A survey of consumer preferences. In *Transportation Research Board 96th Annual Meeting, Washington, DC*, pages 8–12, 2017.
- [2] Emeli Adell. *Driver experience and acceptance of driver support systems-a case of speed adaptation*. Department of Technology and Society, Lund University, 2009.
- [3] Emeli Adell. Acceptance of driver support systems. In *Proceedings of the European conference on human centred design for intelligent transport systems*, volume 2, pages 475–486. Humanist VCE Berlin, Germany, 2010.
- [4] National Highway Traffic Safety Administration. *Investigation Report PE 16-007*. U.S. Department of Transportation, Washington D.C, 2017.
- [5] National Highway Traffic Safety Administration. *Traffic Safety Facts, 2016 Data*. U.S. Department of Transportation, Washington D.C, 2018.
- [6] National Highway Traffic Safety Administration. *Traffic Safety Facts, 2017 Data, Large Trucks*. U.S. Department of Transportation, Washington D.C, 2019.
- [7] Lisanne Bainbridge. Ironies of automation. In *Analysis, Design and Evaluation of Man-Machine Systems 1982*, pages 129–135. Elsevier, 1983.
- [8] Matthias Beggiano, Franziska Hartwich, Katja Schleinitz, Josef Krems, Ina Othersen, and Ina Petermann-Stock. What would drivers like to know during automated driving? information needs at different levels of automation. In *7. Tagung Fahrerassistenzsysteme*, 2015.
- [9] E Bekiaris, S Petica, and K Brookhuis. Driver needs and public acceptance regarding telematic in-vehicle emergency control aids. In *Mobility for everyone, 4th world congress on intelligent transport systems*, 1997.
- [10] Marek Bekier, Brett RC Molesworth, and Ann Williamson. Tipping point: The narrow path between automation acceptance and rejection in air traffic management. *Safety science*, 50(2):259–265, 2012.
- [11] Mathias Benedek and Christian Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1):80 – 91, 2010.
- [12] Wonshik Chee and Masayoshi Tomizuka. Vehicle lane change maneuver in automated highway systems. Technical report, Institute of Transportation Studies, UC Berkeley, 1994.
- [13] Fred D Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340, 1989.
- [14] Fred D Davis, Richard P Bagozzi, and Paul R Warshaw. User acceptance of computer technology: a comparison of two theoretical models. *Management science*, 35(8):982–1003, 1989.
- [15] AP De Vos and W Hoekstra. Behavioural aspects of automatic vehicle guidance (avg); leaving the automated lane. 1997.
- [16] Dick de Waard. *The measurement of drivers' mental workload*. Groningen University, Traffic Research Center Netherlands, 1996. ISBN 9789068073089.
- [17] Birsen Donmez, Linda Ng Boyle, John D Lee, and Daniel V McGehee. Drivers' attitudes toward imperfect distraction mitigation strategies. *Transportation research part F: traffic psychology and behaviour*, 9(6): 387–398, 2006.

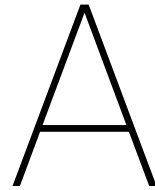
- [18] Donald R Drew, Lynn R LaMotte, JH Buhr, and JA Wattleworth. Gap acceptance in the freeway merging process. *Highway research record*, 208:1–36, 1967.
- [19] Kim Drnec and Jason S. Metcalfe. Paradigm development for identifying and validating indicators of trust in automation in the operational environment of human automation integration. In *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience*, pages 157–167. Springer International Publishing, 2016.
- [20] Mary T Dzindolet, Linda G Pierce, Hall P Beck, and Lloyd A Dawe. The perceived utility of human and automated aids in a visual detection task. *Human Factors*, 44(1):79–94, 2002.
- [21] Mary T Dzindolet, Scott A Peterson, Regina A Pomranky, Linda G Pierce, and Hall P Beck. The role of trust in automation reliance. *International journal of human-computer studies*, 58(6):697–718, 2003.
- [22] Tobias Ehlgen and Tomas Pajdla. Maneuvering aid for large vehicle using omnidirectional cameras. In *Applications of Computer Vision. IEEE Workshop on*, pages 17–17. IEEE, 2007.
- [23] Robertson H.D. Toth G.R. Eicher, J.P. A report to congress on large-truck accident causation. *National Technical Information Service*, 1982.
- [24] Söenke Eilers, Jonas Mårtensson, Henrik Pettersson, Marcos Pillado, David Gallegos, Marta Tobar, Karl Henrik Johansson, Xiaoliang Ma, Thomas Friedrichs, Shadan Sadeghian Borojeni, et al. Companion–towards co-operative platoon management of heavy-duty vehicles. In *Intelligent Transportation Systems (ITSC), 18th International Conference on*, pages 1267–1273. IEEE, 2015.
- [25] Leonard Evans. *Traffic safety and the driver*. Science Serving Society, 1991.
- [26] A. Field. *Discovering Statistics Using IBM SPSS Statistics*. SAGE publications Ltd, 4 edition, 2013.
- [27] Jon Fingas. Waymo launches its first commercial self-driving car service, Dec 2018. URL [www.engadget.com/2018/12/05/waymo-one-launches/](http://www.engadget.com/2018/12/05/waymo-one-launches/).
- [28] Yannick Forster, Frederik Naujoks, and Alexandra Neukum. Increasing anthropomorphism and trust in automated driving functions by adding speech output. In *Intelligent Vehicles Symposium (IV)*, pages 365–372. IEEE, 2017.
- [29] A. Freedy, E. DeVisser, G. Weltman, and N. Coeyman. Measurement of trust in human-robot collaboration. In *International Symposium on Collaborative Technologies and Systems*, pages 106–114, 2007.
- [30] Thomas Friedrichs, Marie-Christin Ostendorp, and Andreas Lüdtkke. Supporting drivers in truck platooning: Development and evaluation of two novel human-machine interfaces. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 277–284. ACM, 2016.
- [31] Alexander Gaedke, Roland Greul, Stefan Kanngießner, and Nicolas Boos. Driver assistance for trucks – from lane keeping assistance to smart truck maneuvering. In *6th International Munich Chassis Symposium*, pages 215–232, Wiesbaden, 2015. Springer Fachmedien Wiesbaden.
- [32] Mahtab Ghazizadeh, John D Lee, and Linda Ng Boyle. Extending the technology acceptance model to assess automation. *Cognition, Technology & Work*, 14(1):39–49, 2012.
- [33] Jordan Golson. Mercedes will give tesla’s autopilot its first real competition this year, Jan 2017. URL [www.theverge.com/ces/2017/1/6/14177872/mercedes-benz-drive-pilot-self-driving-tesla-autopilot-ces-2017](http://www.theverge.com/ces/2017/1/6/14177872/mercedes-benz-drive-pilot-self-driving-tesla-autopilot-ces-2017).
- [34] Adriaan Heino, Hugo Henk Molen, and Gerrit Johan Steven Wilde. *Risk-homeostatic processes in car following behaviour: electrodermal responses and verbal risk estimates as indicators of the perceived level of risk during a car-driving task*. Verkeerskundig Studiecentrum, Rijksuniversiteit Groningen, 1990. ISBN 9789068072105.

- [35] Tove Helldin, Göran Falkman, Maria Riveiro, and Staffan Davidsson. Presenting system uncertainty in automotive uis for supporting trust calibration in autonomous driving. In *Proceedings of the 5th international conference on automotive user interfaces and interactive vehicular applications*, pages 210–217. ACM, 2013.
- [36] John D Hill and Linda Ng Boyle. Driver stress as influenced by driving maneuvers and roadway conditions. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10(3):177–186, 2007.
- [37] B Hills. Some studies of movement, perception, age and accident. In *Publication of: Int'l Congress on Vision and Road Safety, First*. TRRL, 1975.
- [38] M Hoedemaeker. Behoeften, rijstijlen en meningen ten aanzien van automatische voertuig besturing (needs, driving styles and opinions towards automated driving). *Technical University Delft, Delft, The Netherlands. Report*, 1996.
- [39] Bill Howard. What are car surround view cameras, and why are they better than they need to be?, Jul 2014. URL [www.extremetech.com/extreme/186160-what-are-surround-view-cameras-and-why-are-they-better-than-they-need-to-be](http://www.extremetech.com/extreme/186160-what-are-surround-view-cameras-and-why-are-they-better-than-they-need-to-be).
- [40] David C. Howell. *Statistical methods for psychology*. Duxbury Press, 5 edition, 2002.
- [41] Imotions. Galvanic skin response (gsr): The complete pocket guide, Mar 2016. URL [www.imotions.com/blog/galvanic-skin-response](http://www.imotions.com/blog/galvanic-skin-response).
- [42] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4(1):53–71, 2000.
- [43] Alexander Kanaris, Elias B Kosmatopoulos, and Petros A Loannou. Strategies and spacing requirements for lane changing and merging in automated highway systems. *IEEE transactions on vehicular technology*, 50(6):1568–1581, 2001.
- [44] Alexandra Kondyli and Lily Elefteriadou. Driver behavior at freeway-ramp merging areas: Focus group findings. *Transportation Research Record*, 2124(1):157–166, 2009.
- [45] Milos Kyriakidis, Riender Happee, and Joost CF de Winter. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation research part F: traffic psychology and behaviour*, 32:127–140, 2015.
- [46] John D Lee and Neville Moray. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10):1243–1270, 1992.
- [47] John D Lee and Neville Moray. Trust, self-confidence, and operators' adaptation to automation. *International journal of human-computer studies*, 40(1):153–184, 1994.
- [48] John D Lee and Katrina A See. Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1):50–80, 2004.
- [49] Julian Lim, Wen-chau Wu, Jiongjong Wang, John A Detre, David F Dinges, and Hengyi Rao. Imaging brain fatigue from sustained mental workload: an asl perfusion study of the time-on-task effect. *Neuroimage*, 49(4):3426–3435, 2010.
- [50] Abhijai Miglani. User interface design for highly and fully automated driving: Designing for system trust and comfort in non-driving related tasks. Master's thesis, TU Eindhoven and Coventry University, 08 2017.
- [51] Bonnie M Muir and Neville Moray. Trust in automation. part ii. experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39(3):429–460, 1996.
- [52] United Nations. Vienna convention on road traffic, Nov 1968.
- [53] Council of European Union. Council regulation (EU) no 235/59, 1996.

- [54] Council of European Union. Directive 2003/88/ec of the european parliament and of the council concerning certain aspects of the organisation of working time, 2003.
- [55] International Council on Clean Transportation. European vehicle market statistics: Pocketbook 2017/18, 2017.
- [56] Bahar Öz, Türker Özkan, and Timo Lajunen. Professional and non-professional drivers' stress reactions and risky driving. *Transportation research part F: traffic psychology and behaviour*, 13(1):32–40, 2010.
- [57] Raja Parasuraman and Victor Riley. Humans and automation: Use, misuse, disuse, abuse. *Human factors*, 39(2):230–253, 1997.
- [58] Jonathan W Peirce. Psychopy—psychophysics software in python. *Journal of neuroscience methods*, 162(1-2):8–13, 2007.
- [59] Jonathan W Peirce. Generating stimuli for neuroscience using psychopy. *Frontiers in neuroinformatics*, 2:10, 2009.
- [60] AR Quimby and GR Watts. Human factors and driving performance. Technical report, 1981.
- [61] Nadine B Sarter, David D Woods, and Charles E Billings. Automation surprises. *Handbook of human factors and ergonomics*, 2:1926–1943, 1997.
- [62] Takaaki Sato, Alessandro Moro, Atsushi Sugahara, Tsuyoshi Tasaki, Atsushi Yamashita, and Hajime Asama. Spatio-temporal bird's-eye view images using multiple fish-eye cameras. In *System Integration (SII), International Symposium on*, pages 753–758. IEEE, 2013.
- [63] Thomas B. Sheridan. *Telerobotics, Automation, and Human Supervisory Control*. The MIT Press, 1992. ISBN 9780262193160.
- [64] A Smiley and KA Brookhuis. Alcohol, drugs and traffic safety. road users and traffic safety. *Publication of: van Gorcum & comp BV*, 1987.
- [65] Bryant Walker Smith. Human error as a cause of vehicle crashes. *Center for Internet and Society*, 2013.
- [66] Bryant Walker Smith. Sae levels of driving automation. *Center for Internet and Society, Stanford Law School*, 2013.
- [67] Masoud Tabibi. Design and control of automated truck traffic at motorway ramps. 2004.
- [68] DH Taylor. Drivers' galvanic skin response and the risk of accident. *Ergonomics*, 7(4):439–451, 1964.
- [69] The Tesla Team. Introducing navigate on autopilot, Oct 2018. URL [www.tesla.com/nl\\_NL/blog/introducing-navigate-autopilot?redirect=no](http://www.tesla.com/nl_NL/blog/introducing-navigate-autopilot?redirect=no).
- [70] Mercedes-Benz Trucks. Mercedes-benz trucks: Greater safety, 2019. URL [www.mercedes-benz-trucks.com/en\\_GB/models/new-actros/efficiency/greater-safety.html](http://www.mercedes-benz-trucks.com/en_GB/models/new-actros/efficiency/greater-safety.html).
- [71] Volvo Trucks. Volvo driver support systems, 2019. URL [www.volvotrucks.com/en-lb/trucks/volvo-fh-series/features/driver-support-systems.html](http://www.volvotrucks.com/en-lb/trucks/volvo-fh-series/features/driver-support-systems.html).
- [72] Robert van den Brink, Luuk Brederode, and Menno Wagenaar. *Onderzoek naar de wegtype-verdeling en samenstelling van het wegverkeer Eindrapport*. 2010.
- [73] Jinke D Van Der Laan, Adriaan Heino, and Dick De Waard. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation research. Part C, Emerging technologies*, 5(1):1–10, 1997.
- [74] Ministerie van Infrastructuur en Milieu. Reglement verkeersregels en verkeerstekens 1990, 1990.
- [75] J. Vos. *Richtlijn Ontwerp Autosnelwegen 2017*. Ministerie van Infrastructuur en Milieu, Rijkswaterstaat Grote Projecten en Onderhoud (RWS, GPO), m.m.v. Witteveen Bos, 2017.

- 
- [76] Guy H Walker, Neville A Stanton, Tara A Kazi, Paul M Salmon, and Daniel P Jenkins. Does advanced driver training improve situational awareness? *Applied ergonomics*, 40(4):678–687, 2009.
- [77] Adam Waytz, Joy Heafner, and Nicholas Epley. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, 52:113–117, 2014.
- [78] Guangchuan Yang, Hao Xu, Zhongren Wang, and Zong Tian. Truck acceleration behavior study and acceleration lane length recommendations for metered on-ramps. *International journal of transportation science and technology*, 5(2):93–102, 2016.
- [79] Fan Ye and Yunlong Zhang. Vehicle type-specific headway analysis using freeway traffic data. *Transportation Research Record*, 2124(1):222–230, 2009.
- [80] Mark S Young, Stewart A Birrell, and Neville A Stanton. Safe driving in a green world: A review of driver performance benchmarks and technologies to support ‘smart’ driving. *Applied ergonomics*, 42(4):533–539, 2011.
- [81] Ding Zhao, Huei Peng, Kazutoshi Nobukawa, Shan Bao, David J LeBlanc, and Christopher S Pan. Analysis of mandatory and discretionary lane change behaviors for heavy trucks. *Ann Arbor*, 1001:48109–2133, 2017.





# Participant instruction form

## Information for Participants

Study: Automated merging

Dear Sir or Madam,  
thank you for the interest in my research study. Please read the following information carefully and decide whether you want or do not want to participate. Both participating and not participating is free to you. If you have further questions regarding the study, I will be glad to answer them.

### 1. Purpose of this Research

This study investigates truck driver behaviour while merging onto the freeway. We are interested in the effects of presenting visual information to drivers while automated merging on decision making, workload and acceptance.

### 2. Study Procedure

You will be sitting in TU Delfts 'truck simulator'. Your task will be to monitor the situation while the truck is merging onto freeway. The truck will adhere to the traffic rules of the Netherlands. You will get feedback about the automations status. After each merge the simulation will be restarted. At the end of each interface you will answer a questionnaire. During the experiment two electrodes will be placed on your left hand to measure skin conductance.

**Please fulfil the tasks as conscientious and careful as possible.**

### 3. Advantages

#### *a) for Participants*

Your advantage is to get insights in current truck research and to help design driver assistance systems.

#### *b) for Research*

Your data will be used to design new driver assistance systems, hence you make a contribution to the future design of advanced driver assistance systems (ADAS). The study will also give insights to decision making processes in traffic, perceived workload and the acceptance of future ADAS to enhance road safety.

### 4. Risks (Side effects, Inconveniences) for Participants

There are no risks for you.

Mild skin irritation is possible from the electrodes

### 5. Obligation of participants

By means of consent to the participation you undertake *to read the instructions carefully and to follow the instructions conscientiously* and fulfil the given task thoroughly.

### 6. Requirements for participants

Minimum 4 years in possession of a (truck) driving licence, Age >18 Years, normal or corrected to normal vision.

### 7. Confidentiality and handling of the data

The recorded data will be safely stored and is confidential. To guarantee the anonymity of your data, your name is stored separately. Merely, your demographic data will be stored anonymously within the dataset. The examiner is oblique to discretion.

Figure A.1: Participants instruction form - page 1

**8. Voluntary participation**

Your participation is voluntarily. You will not have any disadvantages if you do not participate. You can quit the voluntary participation at any time during the study without the indication of reasons. Also the examiner can make the decision to stop the whole experiment or to stop your participation prematurely if necessary (i.e. for medical reasons).

**9. Confidentiality**

The assessment of the personal data is fully anonymised. No one is able to track your personal data to the experimental data. The anonymised data is used in the experimental analysis and is used to publish results in scientific journals.

**10. Responsible staff**

Felix Dreger

Department of Cognitive Robotics

Intelligent vehicles group

Mekelweg 2

2628CD Delft

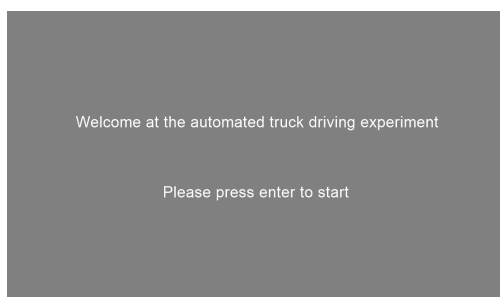
F.A.Dreger@tudelft.nl

Figure A.2: Participants instruction form - page 2

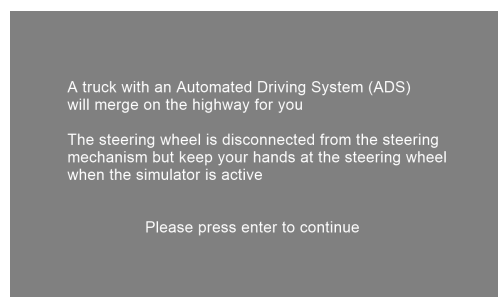


# B

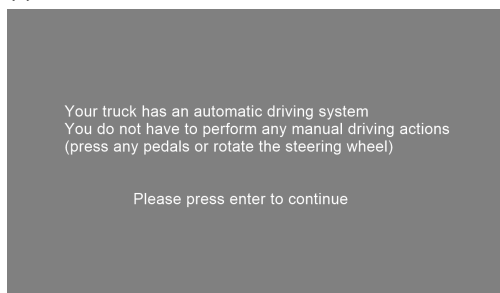
## Simulator screenshots



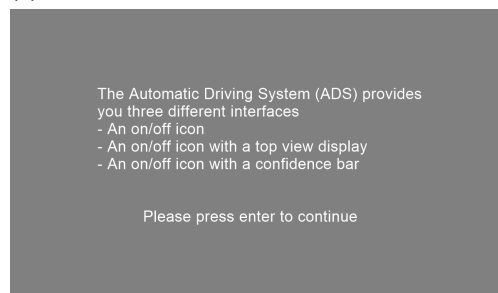
(a) Screen 1



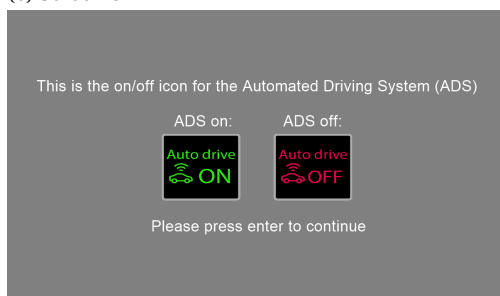
(b) Screen 2



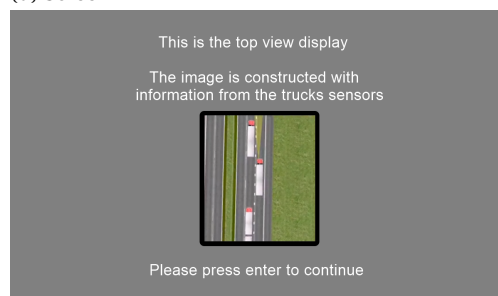
(c) Screen 3



(d) Screen 4

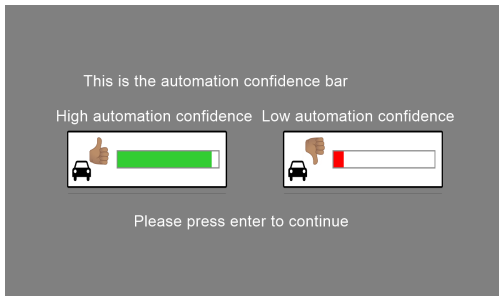


(e) Screen 5

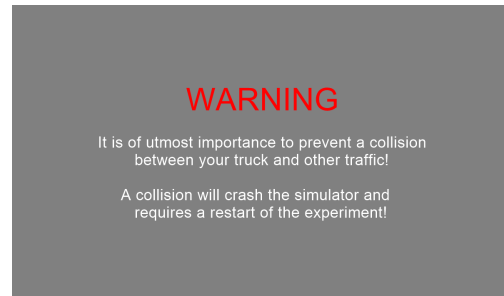


(f) Screen 6

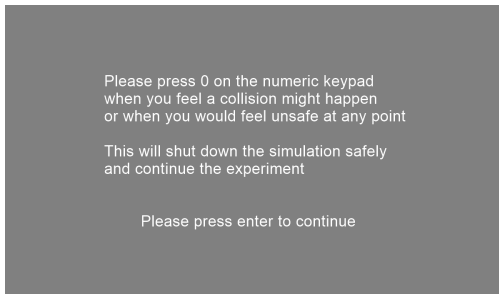
Figure B.1: Simulator instructions



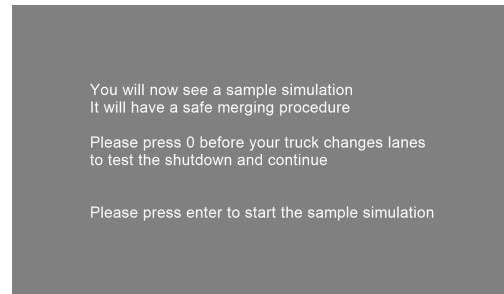
(a) Screen 7



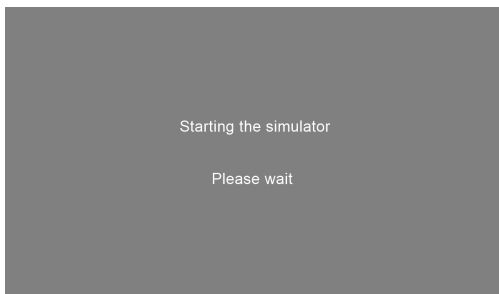
(b) Screen 8



(c) Screen 9



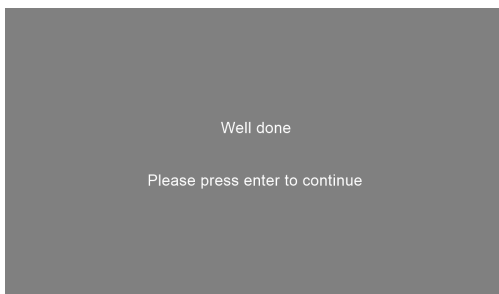
(d) Screen 10



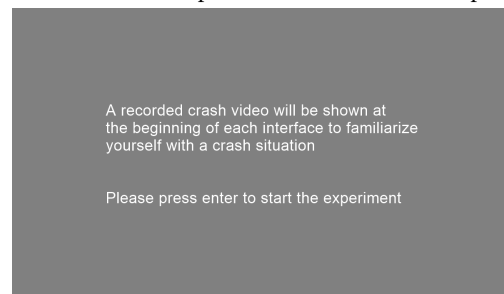
(e) Screen 11



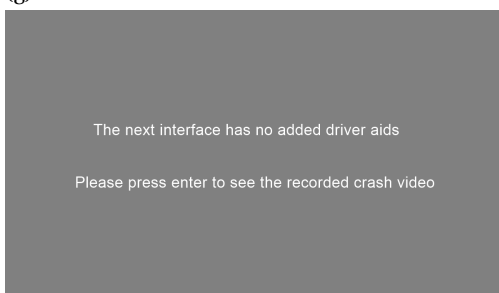
(f) Screen 12 - sample simulation for abortion practice



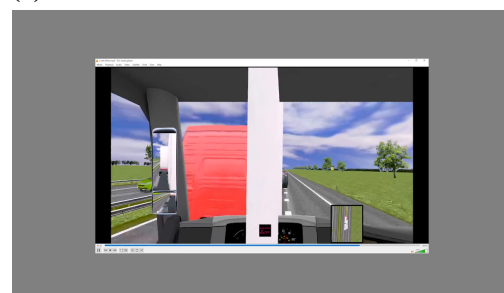
(g) Screen 13



(h) Screen 14

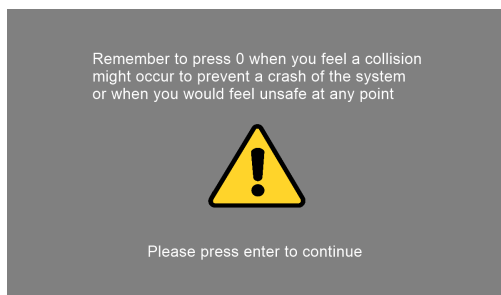


(i) Screen 15

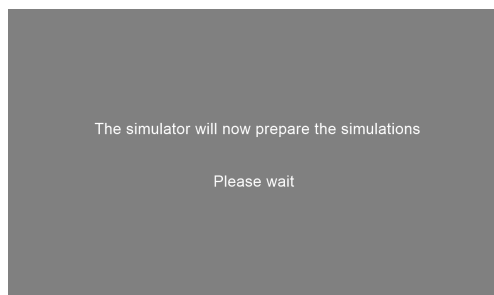


(j) Screen 16 - crash video

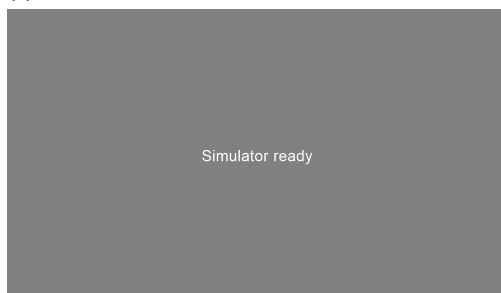
Figure B.2: Simulator instructions



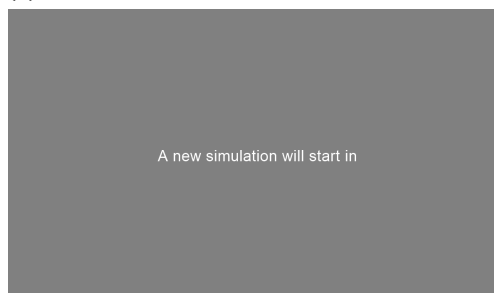
(a) Screen 17



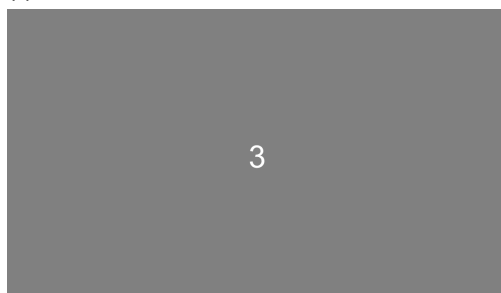
(b) Screen 18



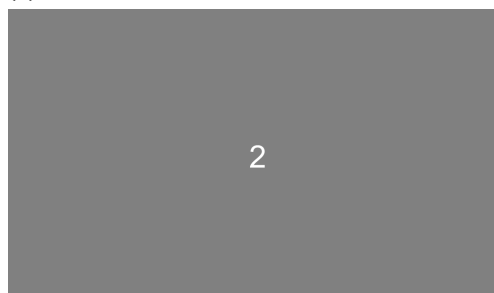
(c) Screen 19



(d) Screen 20



(e) Screen 21



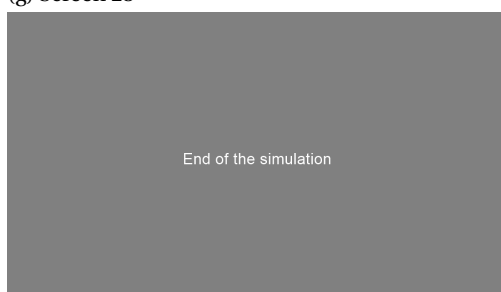
(f) Screen 22



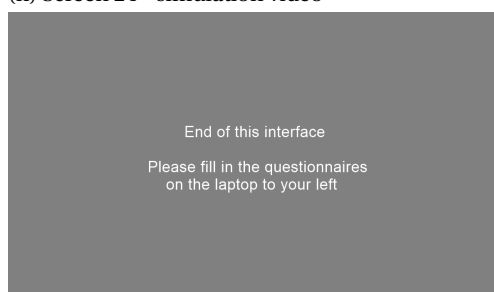
(g) Screen 23



(h) Screen 24 - simulation video



(i) Screen 25



(j) Screen 26

Figure B.3: Simulator instructions



# C

## Experiment questionnaires

### C.1. Demographic questions

Sex

Male

Female

Age

Korte antwoordtekst .....

Number of driver license years

Korte antwoordtekst .....

How often have you used 'Adaptive Cruise Control' while driving? \*

1 2 3 4 5 6

Never       Always

How often have you used 'Automatic Lane Keeping' while driving? \*

1 2 3 4 5 6

Never       Always

Figure C.1: Demographic questions before the start of the experiment

### C.2. Van der Laan's acceptance questionnaire

---

My judgements of the (...) system are... (please tick a box on every line)

---

1	useful	_____	useless
2	pleasant	_____	unpleasant
3	bad	_____	good
4	nice	_____	annoying
5	effective	_____	superfluous
6	irritating	_____	likeable
7	assisting	_____	worthless
8	undesirable	_____	desirable
9	raising alertness	_____	sleep-inducing

---

Figure C.2: Van der Laan's system acceptance questionnaire [73]

### C.3. Jian's trust questionnaire

(Note: not at all=1; extremely=7)

1	The system is deceptive	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	The system behaves in an underhanded manner	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	I am suspicious of the system's intent, action, or outputs	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	I am wary of the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	The system's actions will have a harmful or injurious outcome	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	I am confident in the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	The system provides security	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	The system has integrity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	The system is dependable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	The system is reliable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	I can trust the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12	I am familiar with the system	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure C.3: Jian's trust questionnaire [42]