The influence of take-over requests on driver workload: The role of personality

A driving simulation self-experiment

T. Marfoglia

October 2020

TUDelft

THE INFLUENCE OF TAKE-OVERS ON DRIVER WORKLOAD: THE ROLE OF PERSONALITY

A driving simulation self-experiment

In partial fulfilment of the requirements for the degree of

Master of Science
in **Transport, Infrastructure and Logistics**

in partial fulfilment of the requirements for the annotation
Integral Design and Management

at the Delft University of Technology,

to be defended publicly on Monday November 2, 2020 at 14:45.

by

T. (Themis) Marfoglia

Student number: 4713486

Thesis committee

Chair	Prof. dr. M.P. Hagenzieker	Delft University of Technology
Supervisor	Dr. D.D. Heikoop	Delft University of Technology
Supervisor	Dr. ir. J.C.F. de Winter	Delft University of Technology

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

DOI: 10.4121/13102763.

This study is funded through the NWO-sponsored project, called Meaningful Human Control over Automated Driving Systems [Project Number = 313-99-329].

ACKNOWLEDGEMENTS

In front of you lies a thesis which was written during a very tumultuous period in the midst of a global pandemic. This brought quite a few challenges for the execution of my thesis, however, the resilience of my thesis committee allowed me to successfully complete my thesis. I would therefore like to take this opportunity to express my gratitude to them and everyone who supported me during my research.

First of all, I would like to thank my daily supervisor Dr. Daniël Heikoop for his enormous contribution in the various phases of my research. His enthusiasm during the research is really something to look back on. Thank you for all our help along the way, your guidance helped me enormously during my research.

Thanks to Dr. Ir. Joost de Winter, his critical look at the research was invaluable and contributed significantly to this research.

I would also like to thank to Prof. Dr. Marjan Hagenzieker, for her care and kindness during all the meetings.

Special thanks to Paul van Gent for his continuous support throughout my thesis. Thanks for your help in developing the driving simulator experiment and for lending me your equipment. Your help went to great lengths.

I would also like to thank Solmaz Razmi Rad, she was always available for any questions during the design of the driving simulator experiment.

Special thanks to my family and boyfriend, Rick, for their love and continuous support during all the ups and downs in writing this thesis during a pandemic. This would not have been possible without you. Thanks to my friends, who where there for me and made this journey unforgettable.

Themis Marfoglia
The Hague, October 2020

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACC Adaptive Cruise Control
ADS Automated Driving System
ADS-DV Automated Driving System-Dedicated Vehicle
BFI Big Five Inventory
bpm beats per minute
DDT Dynamic Driving Task
ECG Electrocardiogram
GFP General Factor of Personality
GSR Galvanic Skin Response
HF High-Frequency
HMI Human-Machine-Interaction
HR Heart Rate
HRV Heart Rate Variability
IBI Inter-Beat Interval
LF Low-Frequency
MDSI Multidimensional Driving Style Inventory
MHC-ADS Meaningful Human Control over Automated Driving Systems
NDRT Non-Driving Related Task
NNI Normal-to-Normal Interval
ODD Operational Design Domain
OEDR Object and Event Detection and Response
PPG Photoplethysmography
PPI Peak-to-Peak Interval
RMSSD Root Mean Square of Successive Differences
RRI R-to-R Interval
RTLX Raw Task Load Index
SAE Society of Automotive Engineers
SCL Skin Conductance Level
SD Standard Deviation
SDLP Standard Deviation of Lateral Position
SDNN Standard Deviation of Normal to Normal intervals
TLX Task Load Index
TOR Take-Over Request
TOrt Take-Over reaction time
TTC Time-To-Collision
ULF Ultra-Low-Frequency

EXECUTIVE SUMMARY

Introduction

The development of automated vehicles on the road is in full swing. As vehicles are getting increasingly automated, the human factor is diminished or eventually removed from automated driving. Until then, a combination of human input and automation is necessary during automated driving. This research focuses on the interaction between humans and machine and how a safe interaction can be designed by incorporating meaningful human control. This could be achieved by including individual differences in manual and automated driving in the design of automated vehicles. Previous studies have already shown that a link exists between personality and manual driving behaviour. It is yet unknown how personality is expressed during take-overs, which is critical for save driving in automated vehicles. This research intended to investigate whether personality plays a role in driver workload, as workload plays a vital role in take-over performance. To this end, the following research question was formulated:

To what extent does personality interact with driver workload induced by a take-over request (TOR)?

Methodology

A driving simulation experiment was designed to investigate the role of personality in driver workload at take-overs. However, due to COVID-19, the aim shifted to validating the design and research set-up of the aforementioned experiment. This study therefore provides an empirically validated set of design variables for the study regarding the role of personality traits in automated driving. An $N = 1$ experiment was performed with the researcher as the sole participant. Design variables that are found to play a role in driver workload were varied in the validation experiment. These variables are the duration of the time budget (0, 5, 10 and 15 seconds), traffic density (0, 5, 10 and 15 vehicles/km/lane), location of the TOR (straight stretch and curve) and task during automated driving (monitoring or playing Tetris). In total, the experiment included 64 scenarios, all unique in their characteristics.

Subsequently, workload was measured by a combination of subjective and physiological indicators and driving performance. The Raw Task Load Index (RTLX), a simplification of the NASA Task Load Index (TLX), was used for subjective workload. Heart Rate (HR) and Heart Rate Variability (HRV) measures were included for physiological workload, and driving performance was indicated by the take-over reaction time (TOrt), and Standard Deviation (SD) of speed and Standard Deviation of Lateral Position (SDLP) after the take-over. Notably, this study includes the Root Mean Square of Successive Differences (RMSSD) and Standard Deviation of Normal to Normal intervals (SDNN) as HRV measures, which is a novel approach in studies measuring TOR-induced workload. It was intended to use the Tetris scores obtained during automated driving as secondary task workload indicator. Tetris, therefore, would be both a design variable (one of the varied tasks during automated driving) and a workload measure. However, as the scores did not reflect driver workload, it was not used as workload measure.

Discussion of the results

The driving simulation was not experienced as very demanding, but it could also not be considered an easy or relaxing activity. As the experiment required the participant to repeatedly drive similar scenarios in a short time span, developments in perceived and experienced workload were measurable. As such, it was found that using the RTLX possibly led to respondent fatigue and response bias, a known drawback when using questionnaires in experiments. Moreover, physiological workload decreased as the experiment progressed, which suggests a time-on-task effect. Also, TOR-induced workload decreased as the experiment progressed, which can be attributed to the participant's improved ability to anticipate the TOR. Furthermore, prerequisite knowledge of the researcher about the simulation possibly resulted in high driving performance measurements.

In general, the various workload measures did not unambiguously indicate differences in workload for the various design variables. As such, for a longer duration of the time budget, decreases in workload were measured by the RTLX, although the SDNN and SDLP suggested that workload only decreased when the time budget increased from 0 to 5 seconds. For increases in traffic density, driving performance suggested that workload increased as well. Although, subjective and physiological workload suggested that TOR-induced workload increased with traffic density increasing up to 10 vehicles/km/lane, but decreased at a high traffic density of 15 vehicles/km/lane. Regarding the TOR location, the workload measures all indicated that a TOR in a curve induced more workload than a TOR on a straight stretch of the road. Moreover, the two tasks during automated driving led to similar TOR-induced workload, although emerging out-of-the-loop issues were expected to result in different effects on workload.

Conclusions

Unfortunately, due to the COVID-19 circumstances the main research question remains open for the future study with over 100 participants regarding personality in automated driving. However, based on the results of the $N = 1$ experiment recommendations are made to include specific design variables in the future personality experiment, see Table 0.1. It is recommended to simulate TORs at both low and high demand, as it is expected that people differing in personality exhibit different behaviour at different task demands.

Table 0.1: Recommendations on design variables in future experiments

Design variables	Recommendations
Time budget [s]	Low demand: 10- or 15-second time budget High demand: 0-second time budget
Traffic density [veh/km/lane]	Low demand: 0 vehicles/km/lane High demand: 5, 10 or 15 vehicles/km/lane
Location of the TOR	Low demand: straight road High demand: curve
Task during automated driving	Short duration: monitoring Long duration: monitoring or task with a constant demand

To measure workload, it is recommended to use a similar methodology as in the current study, with a combination of subjective, physiological, driving performance and secondary task measurements. The results of this study are not conclusive as to whether the RMSSD and SDNN are sufficiently sensitive to variations in task demand. Future studies are advised to include these measures to obtain a better understanding of the validity of the RMSSD and SDNN as workload measure. Furthermore, it is recommended to use a different TOrt calculation method and to use a different modality to manually resume control. As secondary task performance measure, it is recommended to the use the proven n-back task instead of using the Tetris scores.

Little speculation can be made about the effect of take-overs on workload in a future study involving participants differing in personality traits. However, reference values are provided for the workload measurements, although the values should be taken with great caution as they are based on a $N = 1$ study. In addition, it is recommended to simulate only a limited number of experiment runs per person, in a counterbalanced order, to limit learning effects and time-on-task effects which were found in the current experiment.

All in all, this research has provided recommendations on the personality study, not only limited to the design of the driving simulation, but also how the experiment can be conducted and analysed. All in all, conducting an $N = 1$ study proved valuable for validating the design of the driving simulation experiment, allowing a more focused approach in future studies.

1 INTRODUCTION TO THE STUDY

1.1 PERSPECTIVES ON AUTOMATED DRIVING AND ROAD SAFETY

With 'Europe on the Move', the European Commission set out an ultimate goal to reach a 'Vision Zero' objective of almost zero road deaths by 2050 (European Commission, 2017). To reach this goal, an intermediate goal has been set to halve the number of road deaths in 2020 compared to 2010, and to halve the number of serious injuries by 2030 compared to 2020 (Council of the European Union, 2017). However, in recent years, progress in reducing the number of road accidents and deaths stagnated throughout Europe. Some countries are even experiencing an increase in the number of road accidents and deaths (Adminaite-Fodor et al., 2019).

This high number of traffic accidents and fatalities is caused by multiple factors, of which the human factor has the biggest impact. Research by the National Highway Traffic Safety Administration (2008) showed that in more than 90% of all crashes human factors were a contributory factor or the sole cause. Human factors, in this case, include recognition errors, decision errors, performance errors, and non-performance errors (National Highway Traffic Safety Administration, 2008). Hence, the human factor must be reduced to eventually reach the 'Vision Zero' objective. The European Commission (2017), therefore, urges the need for automated mobility to diminish or fully remove the human factor while driving. However, challenges still need to be solved to ensure that automated vehicles can effectively analyse the environment and act accordingly to ensure safety, better than a human driver. At present, the current fleet therefore mainly consists of vehicles equipped with features that aid the driver in performing the driving task. These features are known as the Automated Driving System (ADS) of a vehicle and provide aid by informing the driver, by issuing warnings in critical driving situations, by taking over the driving task and by intervening when a critical situation appears. Ultimately, no human driver is needed.

The United States currently leads the way in on-road testing of driverless vehicles. For instance, Waymo[1] is testing their self-driving car on public roads in 24 cities across the United States. Since 2018 Waymo began to commercially offer a self-driving ride hailing service in Phoenix (Waymo LLC, 2020). However, these vehicles often get bad publicity when something goes wrong during automated driving and a road user gets injured or even dies as a result of the injuries from the accident. This became global news on March 18, 2018, when for the first time a road user died in a traffic accident involving a driverless vehicle (Levin and Wong, 2018). In Europe, fully automated and driverless vehicles are not yet allowed on public roads, although tests with these vehicles are allowed on a small scale. Before these vehicles are allowed on the European market, the vehicles must be developed and tested further. Until then, the European Commission has set high standards for vehicles equipped with ADS. As a result, European vehicles are restricted in their functionality, for instance, Tesla's 'autopilot' and 'full self-driving' functions are limited to the extent that the driver should always be attentive and keep their hands on the steering wheel, whilst those functions were actually intended to allow the driver to have hands off the steering wheel as often as possible (Tesla, 2020).

To be admitted to the European market, vehicles must comply with the guidelines of the European Commission as adopted on April 9, 2019 (European Commission, 2019). The guidelines are intended for vehicles of Society of Automotive Engineers (SAE) level 3 and 4

1 Formerly known as the Google self-driving car project (Waymo LLC, 2020)

that can drive automated in a limited number of situations, for instance only on the highway. For fully automated and driverless vehicles there are no guidelines yet for admission to the European market, as these vehicles were not expected to be ready for commercial use in 2019. The European Commission agreed in European Commission (2019) to revise the guidelines in 2020 to reflect technological progress, however no revision to the guidelines have been published to date.

1.2 PROBLEM STATEMENT

Driving an automated vehicle is limited to its Operational Design Domain (ODD). Specified in terms of conditions, the ODD of an automated vehicle is limited to drive on specific road conditions, geographic areas, environmental conditions, a speed range, but can also include other conditions to ensure safe automated driving (SAE International, 2018). Different levels of automated driving can be distinguished, of which the best-known categorisation is created by Society of Automotive Engineers (SAE) and is illustrated in Figure 1.1. Six levels are distinguished that range from level 0 without an Automated Driving System (ADS) to level 5 where the ADS is no longer limited by an ODD. The intermediate levels have a limited ODD.

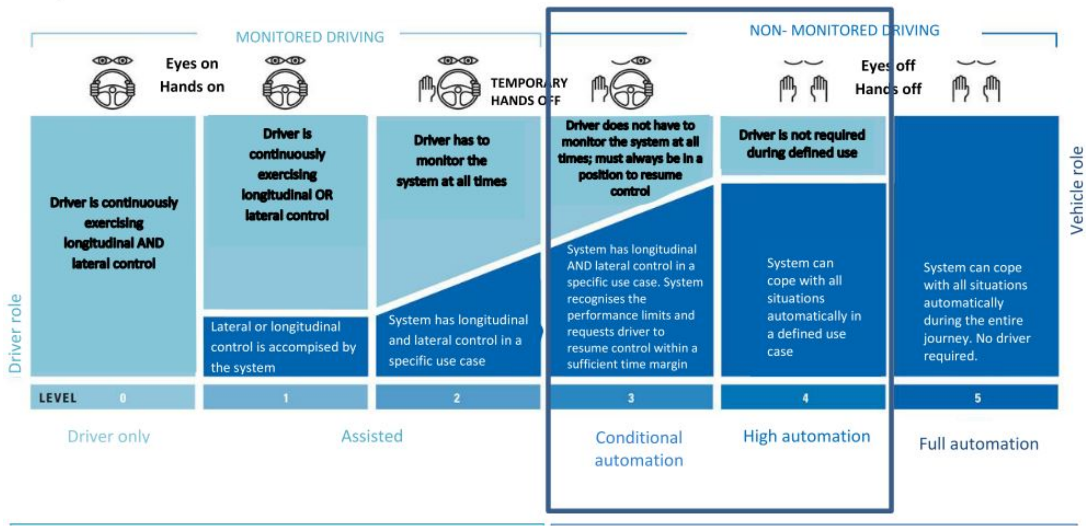


Figure 1.1: Levels of automated driving (European Commission, 2019)

As shown in Figure 1.1, the driver does not need to monitor the ADS in conditional and high automation. The difference between the two levels is that only with conditional automation, the driver must always be available to take over the driving task from the ADS. Vehicles with high automation do not require the driver to be available to take over the driving task during automated driving. In this specified ODD the vehicle can be driverless. When driving outside the specified ODD, the driver could be asked to take over the driving task. In the European Commission (2019) guidelines, it is specified when the driver can be requested to take over the driving task in vehicles with conditional and high automation:

"The system may request the driver to take over with a sufficient lead time in particular when the system determines that it is difficult to continue [the] automated driving mode, such as when the situation becomes outside the [ODD], or when a problem has occurred to the automated vehicle." (European Commission, 2019, p. 4)

The guidelines of the European Commission (2019) also contain a safety requirement about driver availability to the take-over request (TOR):

"If the system is designed to request the driver to take over under some circumstances, the system shall monitor whether the driver is ready to take over driving from the system. It shall ensure through appropriate design (e.g. driver monitoring system) and warnings that the driver remains available to respond to take over request and prevent any foreseeable and preventable misuse by the driver in the [ODD]." (European Commission, 2019, p. 5)

This requirement poses a challenge for designing an ADS that meets the above mentioned standard. However, the actual challenge for designing such ADS goes beyond this requirement, by not only designing a system which is able to ensure driver availability to respond to the TOR but also to ensure that the driver is able to do so in a safe manner. This issue is being studied in the MHC-ADS project, which builds on how meaningful human control can be developed for automated vehicles to achieve a responsible transition to automated driving. A uniform approach may not be suitable to get drivers to take over the driving task from the ADS safely Körber and Bengler (2014). Possibly, individual differences in manual and automated driving can be included in the design of the ADS Heikoop et al. (2019b). Including personality in automated driving is in line with the emergence of tailor-made Human-Machine-Interaction (HMI), which takes into account the complexity of the driving environment and driver's state for transferring control between the driver and the ADS Dukic Willstrand et al. (2018). It is already known that a link exists between personality and manual driving behaviour Taubman-Ben-Ari et al. (2004). For instance, nervousness and anxiousness are attributed to Neuroticism (one of the Big Five Cattell (1957), which is linked to a low tendency of risk-taking traffic behaviour Taubman-Ben-Ari and Yehiel (2012). However, it is unknown how personality is expressed in driver workload and thereby, automated driving behaviour and take-over performance.

Driver workload plays an important role in take-over behaviour and performance, where both underload and overload can be detrimental Young and Stanton (2002). Vehicle automation can result in mental underload when task demand is low, for instance during automated driving on the highway, or mental overload when task demand is high, for instance at the take-over Coughlin et al. (2009); Paxion et al. (2014). Workload depends on both context-dependent factors, such as traffic density at the take-over Gold et al. (2016); Scharfe et al. (2020) or engagement in a non-driving related task during automated driving Merat et al. (2012); Zeeb et al. (2016), and person-dependent factors, such as age and driving experience Wright et al. (2016). Numerous researchers studied how people differ in their response to a TOR and how driver workload plays a role. For instance, the study by Gold et al. (2013) showed that drivers differ in their braking and steering behaviour, but also in their reaction time to a TOR. Driver reaction-time was found to range between 1.9 and 25.7 seconds in non-critical take-over situations depending on task engagement before the TOR (Eriksson and Stanton, 2017). From literature, various causes can be identified for these different response types to a TOR. Wright et al. (2016), for instance, showed that driving experience plays a role in how well the driver is able to take over the driving task after a TOR. Drivers aged between 25 and 59 years with more driving experience than drivers aged between 18 and 22 years are more likely to quickly achieve situation awareness after the TOR (Wright et al., 2016). The study by Körber et al. (2016) found a link between driving experience and driving performance after a take-over: more experienced drivers are able to maintain a higher Time-To-Collision (TTC) as they use the brakes more often and more strongly. In addition, less experienced drivers experience more workload in highly complex driving situations, causing their driving performance to deteriorate more than experienced drivers (Paxion et al., 2014). Several other studies also confirmed the existence of a negative relation between workload and take-over performance (see e.g. Yoon and Ji (2019), who found that take-over time increases when drivers experience more workload). Causes for different take-over behaviour are thus found in context-dependent factors, such as situation complexity, but also in individual or person-dependent factors, such as driving experience, both of which influence the degree of experienced workload that affects take-over performance. A person-specific approach can possibly contribute to a achieve safe transitions of control taking into account the personality of the driver. However, to date it is unknown how personality affects driving behaviour in automated vehicles, and how this translated to driver workload differences. This therefore creates a clear knowledge gap as to whether there is a relationship between personality and driver workload. Answers to this can serve as guideline for designing a safe transfer of the driving task from the ADS to the driver.

1.3 SCOPE

This research focuses on the effect of take-over requests (TORs) on driver workload. Originally, the aim was to investigate whether the personality of a driver plays a role in driver workload because of TORs. Ultimately, to develop meaningful human control for automated vehicles to achieve a responsible transition to automated driving Heikoop et al. (2019b). For this, a driving simulation experiment involving participants was intended to be carried out. However, the focus of this study had to be adjusted, because the measures of Delft University of Technology for the COVID-19 pandemic did not allow a driving simulation experiment involving participants. Within the new focus of the research, a self-experiment is carried out in which the researcher herself is the test subject. By means of a driving simulation self-experiment design variables will be varied that are expected to affect driver workload at take-overs. Ultimately, this study provides an empirically-validated set of design variables affecting driver workload for the study regarding the role of personality traits in automated driving. Moreover, recommendations on how to conduct the future study regarding personality are provided, and hypotheses are formulated about the results which can be expected.

1.4 RESEARCH QUESTIONS

Following the problem statement, the following research question is proposed:

To what extent does personality interact with driver workload induced by a take-over request (TOR)?

In order to answer the research question, multiple sub-questions are proposed:

1. What is a suitable design of a driving simulation experiment in which the effect of a TOR on workload can be measured?
2. How can the interaction between personality and workload be investigated in a driving simulation experiment regarding the effect of a TOR on workload?
3. What effect of take-overs on workload can be expected in a study involving participants, based on the results of this self-experiment?

1.5 READING GUIDE

This study is structured as follows. First, relevant literature on automated driving, driver workload and personality traits are presented in Chapter 2. Based on the literature findings, the methodology for the driving simulation was determined. The used methodology is presented in Chapter 3. The experimental set-up is presented, by elaborating on the various design variables, the apparatus, the design of the simulation, driver workload measurement techniques, and the used analysis methods. Then, in Chapter 4 the results of the experiment are presented regarding workload trends and the effect of the various design variables on driver workload. The results will then be discussed in Chapter 5, here a reflection is given on the effect of take-overs on driver workload, and it is argued why certain workload effects have (not) been found in the experiment. Moreover, recommendations for future studies are formulated. Lastly, Chapter 6 concludes this research by formulating answers to the research question and the several sub-questions.

2 LITERATURE REVIEW

This chapter presents the current literature on automated driving, workload and personality. Section 2.1 first explains what automated driving is. A classification of automated driving levels is presented and the notion of transition of control is introduced. Section 2.2 will then elaborate on the transition of control. Here in Section 2.2.1 the notion of Dynamic Driving Task (DDT) fallback is introduced, which is an important notion in higher levels of automated driving. Following the DDT, the importance of the Section 2.2.2 is elaborated on. Then in Section 2.3, workload is defined and the importance of workload in relation to the take-over process is discussed. In this section, factors that play a role in take-over performance are discussed. Here, the role of personality in driver workload will be discussed.

2.1 CLASSIFICATION OF AUTOMATED DRIVING LEVELS

The levels of automation, from completely manual to fully autonomous and driverless, have been defined by several authorities, of which the classification by the Society of Automotive Engineers (SAE International, 2018) is best known:

LEVEL 0: NO DRIVING AUTOMATION 'The performance by the *driver* of the entire *Dynamic Driving Task (DDT)*, even when enhanced by *active safety systems*'.

LEVEL 1: DRIVER ASSISTANCE 'The *sustained* and *Operational Design Domain (ODD)*-specific execution by a *driving automation system* of either the *lateral* or the *longitudinal vehicle motion control* subtask of the *DDT* (but not both simultaneously) with the expectation that the *driver* performs the remainder of the *DDT*'.

LEVEL 2: PARTIAL DRIVING AUTOMATION 'The *sustained* and *ODD*-specific execution by a *driving automation system* of both the *lateral* and *longitudinal vehicle motion control* subtasks of the *DDT* with the expectation that the *driver* completes the *Object and Event Detection and Response (OEDR)* subtask and *supervises* the *driving automation system*'.

LEVEL 3: CONDITIONAL DRIVING AUTOMATION 'The *sustained* and *ODD*-specific performance by an *Automated Driving System (ADS)* of the entire *DDT* with the expectation that the *DDT fallback-ready user* is *receptive* to *ADS*-issued *requests to intervene*, as well as to *DDT performance-relevant system failures* in other *vehicle* systems, and will respond appropriately'.

LEVEL 4: HIGH DRIVING AUTOMATION 'The *sustained* and *ODD*-specific performance by an *ADS* of the entire *DDT* and *DDT fallback* without any expectation that a *user* will respond to a *request to intervene*'.

LEVEL 5: FULL DRIVING AUTOMATION 'The *sustained* and unconditional (i.e., not *ODD*-specific) performance by an *ADS* of the entire *DDT* and *DDT fallback* without any expectation that a *user* will respond to a *request to intervene*'.

According to this classification, there are 5 levels of driving automation and an extra level 0 where the vehicle is fully manual. Vehicles of all automated driving levels are equipped with an ADS that can partially or fully take-over the driving task. Society of Automotive Engineers (SAE) makes a distinction between a driving automation system (used for level 1 and 2 vehicles) and an ADS (used for level 3 to 5 vehicles).

The role of the user of an automated vehicle differs per level. Up to level 2, the user is always the driver while being supported by a driving automation system, while at higher levels of driving automation, the user is sometimes a driver and sometimes a passenger when the ADS is performing the DDT. Specifically, the user of a level 3 vehicle is a receptive and fallback-ready user, which entails that the user must be receptive to take-over requests (TORs) from the ADS and be receptive to DDT performance-relevant system failures for which the ADS does not request a take-over, i.e. a broken suspension component (SAE International, 2018). Levels 4 and 5 do not require the user to take-over the driving task and become the driver. Whether it is possible to become the driver in vehicles of these automated driving levels depends on whether the vehicle is an Automated Driving System-Dedicated Vehicle (ADS-DV), which is a vehicle that is not necessarily equipped with a user interface, such as a steering wheel or pedals[1]. For level 4 non-ADS-DV vehicles, equipped with a user interface, the user has to become the driver to complete the trip when leaving the specific ODD for fully automated driving. For example, a level 4 vehicle is only able to drive fully automated on the highway in low-density traffic conditions, but once the highway has to be exited, the user must become the driver to complete the trip. In level 5, it is not necessary to take over the DDT to complete the trip, but there is the possibility to become the driver if the user wishes to and if the vehicle is equipped with a user interface.

2.2 TRANSITION OF CONTROL

Transitioning control between the driver and the Automated Driving System (ADS) is an essential part of, especially, driving a level 3 vehicle. The take-over is a specific type of transition of control. Other types of transitions of control are classified by Lu and de Winter (2015) and Lu et al. (2016), and can be found in Figure 2.1. For this categorisation, a distinction is made between whether the user or the ADS requests the transition of control and whether the user or the ADS executes the Dynamic Driving Task (DDT) after the request. If the user initiates the request, a hand-over is initiated, whilst if the ADS initiates the request, a take-over is initiated (Riener et al., 2017).

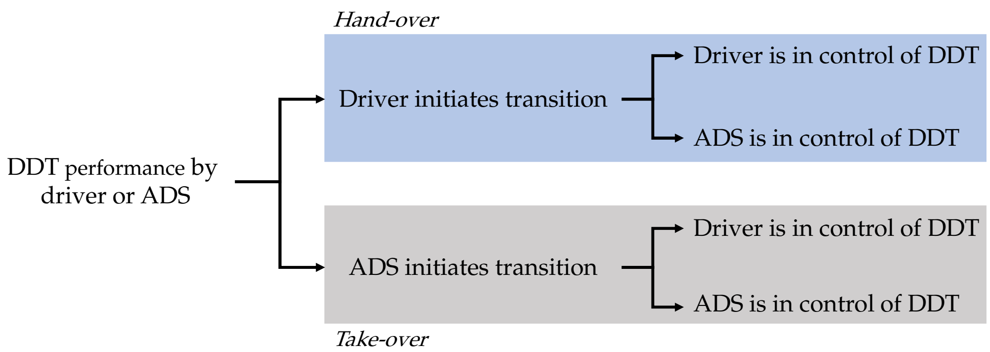


Figure 2.1: Categorisation of transitions of control, based on Lu and de Winter (2015), Lu et al. (2016) and Riener et al. (2017)

1 More information on ADS-DV can be found in SAE International (2018)

As can be seen in Figure 2.1, there are two types of (driver-initiated) hand-overs and (ADS-initiated) take-overs. After a hand-over, either the driver or the ADS is in control of the DDT. For example, the driver can be initiating a transition when noticing a vehicle failure for which the ADS has not issued a take-over request (TOR). The driver could also be initiating a hand-over when entering the Operational Design Domain (ODD) of the ADS, for example when entering the highway (if the ADS does not recognise this itself and requests a take-over). An example of a take-over after which the driver is in control of the DDT is when the ADS requests the fallback-ready user to execute the DDT when approaching the ODD-exit. The ADS could also initiate a take-over after which the ADS itself is in control, for example when it notices the vehicle entering the highway if this is part of the ADS's ODD (and if the driver has not initiated a hand-over already).

Figure 2.2 illustrates the take-over process from automated to manual driving of a level 3 vehicle. Three phases can be distinguished in the take-over process from automated to manual driving (Petermeijer et al., 2016). First, the automated driving phase in which the ADS executes the DDT. Secondly, the transition of control phase in which execution of the DDT is switched between the ADS and the driver. Finally, the manual driving phase in which the driver executes the DDT. Eventually, the driver is back into the loop.

2.2.1 Dynamic Driving Task fallback

Designing safe transitions of control is challenging, especially for level 3 vehicles as the user is expected to be the DDT fallback. A further clarification of the DDT fallback has been provided by SAE International (2018, p. 7):

"At level 3, an ADS is capable of continuing to perform the DDT for at least several seconds after providing the fallback-ready user with a request to intervene. The DDT fallback-ready user is then expected to achieve a minimal risk condition if s/he determines it to be necessary."

Once a TOR is issued, the user is requested to immediately execute the DDT fallback (SAE International, 2018). This does not entail that the ADS is disengaged immediately, the ADS may be able to continue to execute the DDT for a specified time depending on the situation (SAE International, 2018). The time between the TOR and the moment when the ADS is no longer able to perform the DDT is defined as the time budget (Zeeb et al., 2016). Once the user has taken over the DDT and has become the driver, driving can be continued if possible or a minimal risk condition has to be achieved by the driver. An example of such a minimal risk condition is moving the vehicle from the current active lane to the emergency lane.

There are several causes which necessitate the fall-back user to perform the DDT fallback, such as a failure with the ADS or the vehicle itself, or when the exit of the ODD will be reached soon (SAE International, 2018) (Figure 2.2). The ADS requests a take-over for an ADS failure and when approaching the ODD-exit. However, a take-over is not always requested when a vehicle failure occurs. Apparent vehicle failures (e.g. a flat tire) must be noticed by the fallback-ready who must then perform the DDT fallback. The requirement of the fallback-ready user to be receptive to TORs and to DDT performance-relevant system failures for which no TOR is issued makes level 3 vehicles particularly challenging for ensuring driving safety. Especially as users of level 3 vehicles are allowed to engage in Non-Driving Related Tasks (NDRTs) during automated driving, but must also be able to respond to the TOR and to apparent vehicle failures in a timely and safe manner (Inagaki and Sheridan, 2019).

If a driver is not receptive to a TOR, a failure mitigation strategy might be executed by the vehicle if it is equipped with this specific feature (SAE International, 2018). With this feature, the vehicle is able to execute a failure mitigation strategy to achieve a minimal risk condition, as illustrated in Figure 2.2. At present, there is no European regulation yet that requires all vehicles to be equipped with a failure mitigation strategy. Eleven major players in the automated driving industry denounced this lack of regulation in a white paper for standardisation of the automated driving industry (BMW et al., 2019). Three manoeuvres

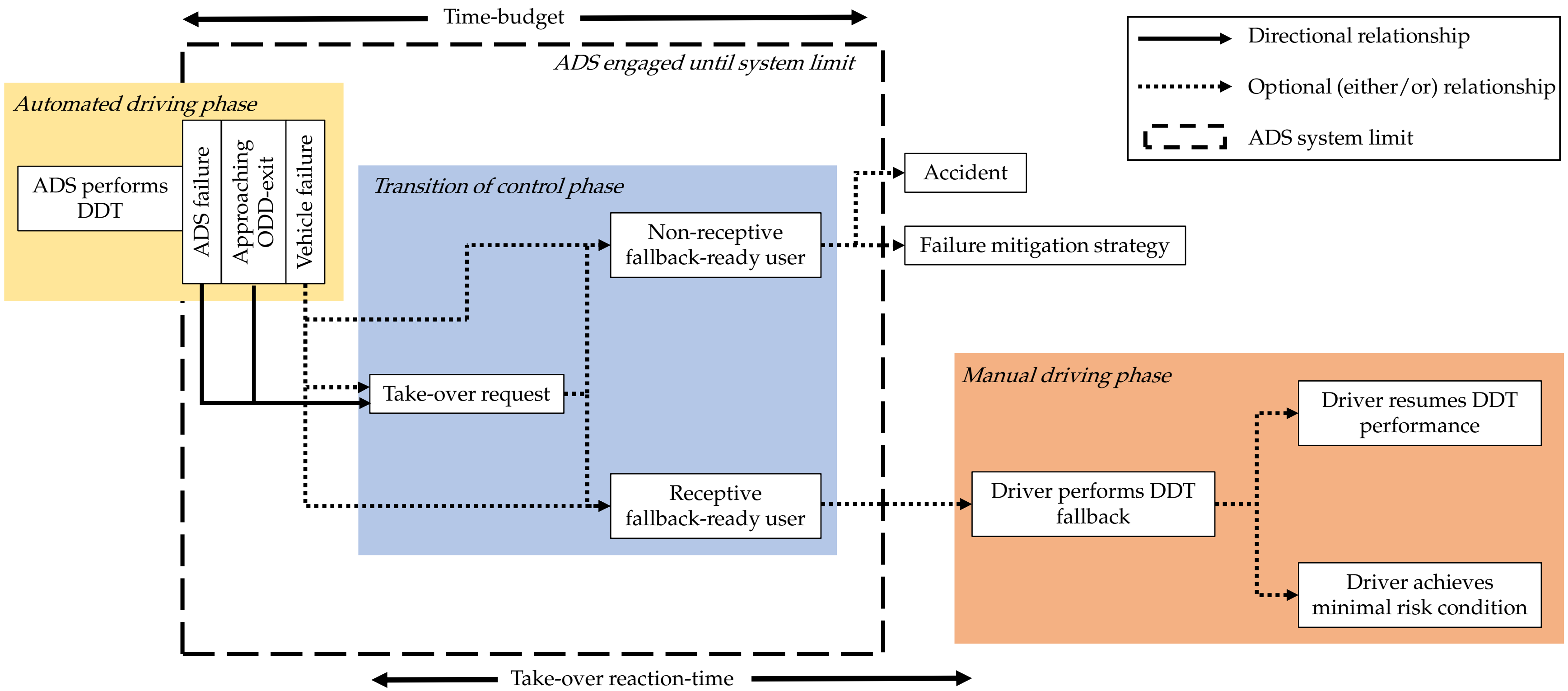


Figure 2.2: Take-over process from automated to manual driving, based on Petermeijer et al. (2016); Marberger et al. (2017) and SAE International (2018)

as part of the failure mitigation strategy were proposed depending on the severity of the ADS failure: a comfort stop, a safe stop and an emergency stop. Tesla vehicles with Autosteer functionality, a feature of level 2 vehicles, deal with a non-receptive driver as proposed in the white paper. When a driver is using Tesla's Autosteer functionality, the driver is required to always keep hands on the steering wheel. If the vehicle does not detect hands on the steering wheel for a certain duration, an auditory warning will be used to ensure that the driver puts the hands back on the wheel. In case of a non-responsive driver, hazard lights will be used to warn surrounding traffic. The vehicle will then proceed to slow down and stop in its current lane (Tesla, 2020).

Level 4 and 5 vehicles are able to perform the fallback and achieve a minimal risk condition. The minimal risk condition that vehicles with a level 4 or 5 ADS are able to perform entails either bringing the vehicle to a stop in its current travel path (but could also be a more extensive manoeuvre that removes the vehicle from its current travel path and moves the vehicle to a non-active or emergency lane) or, if possible, continue driving to a repair facility with adjusted speed and using hazard lights (SAE International, 2018). Some level 4 and 5 vehicles allow the user to perform the fallback and achieve a minimal risk condition (if needed) or continue execution of the DDT. However, the ADS of these vehicles is able to overrule the user in order to reduce crash risk (SAE International, 2018). Level 3 vehicles cannot overrule the user to perform the fallback, even if this would lead to a less safe situation than could be achieved by the failure mitigation strategy (if the vehicle is equipped with this feature).

2.2.2 Take-over time budget

An important element that influences the take-over process, which is depicted in Figure 2.2, is the take-over time budget (also known as take-over lead-time). The time budget is defined as the time from the TOR to the critical event or system limit for which a take-over is requested (Zeeb et al., 2016; Marberger et al., 2017; Eriksson and Stanton, 2017). Three other measurable moments during the take-over process can be distinguished from rthe same figure. These are 'take-over reaction time (TOrt)', 'remaining time budget' and 'exceeded time budget'. The TOrt is defined as the time between the TOR and manual resumption of the DDT. The time between resumption of manual control and reaching the system limit of the ADS, it is either called the 'remaining time budget' when the driver performs the DDT fallback before reaching the system limit (early), whereas it is called the 'exceeded time budget' when the driver performs the DDT after reaching the system limit (late).

Time budget plays a critical role in take-over performance. If the time budget is too limited, the user is allowed less time for cognitive processing and action selection, so lower quality responses are expected (Petermeijer et al., 2016; Gold et al., 2013). Therefore the European Commission (2019, p. 4) had set a guideline regarding the time budget:

- *"The system may request the driver to take over with a sufficient [time budget] in particular when the system determines that it is difficult to continue automated driving mode, such as when the situation becomes outside the [ODD], or when a problem has occurred to the automated vehicle."*

This guideline poses a requirement for providing the driver with a 'sufficient' time budget to ensure a safe take-over. What is sufficient differs per take-over situation (Eriksson and Stanton, 2017). According to Eriksson and Stanton (2017), the time budget must ensure a comfortable take-over for non-critical take-over situations, while for critical situations, comfort is not of importance, as long as the time budget is sufficient to prevent an accident to happen. Thus, time budget must depend on the criticality of the take-over situation. The criticality of the take-over situation cannot be classified based on clear criteria, but it depends on various factors, such as spatial-temporal and road characteristics, but also traffic and weather conditions (Lu et al., 2016). However, regardless of the criticality of the situation, the time budget itself can also provide a greater or lesser sense of urgency. For instance, if a short time budget is provided to the user, it creates a critical take-over in terms of available time for processing the situation and selecting appropriate actions, even if the situation itself is not critical (Van den Beukel and Van der Voort, 2013).

In the same study by Eriksson and Stanton (2017), the authors performed a meta-analysis regarding time budgets in simulated driving studies. It was found that the criticality of the take-over situation does not always go hand in hand with the criticality of the time budgets. They concluded that often in simulated driving studies relatively short time budgets are used for non-critical situations. For example, when approaching a highway construction site, which can be known well in advance due to the vehicle's connection to road information services, the TOR is issued only a few seconds in advance. This creates an unlikely great sense of time-criticality. However, the TOR must not be issued too early, as the reason for the take-over will not be immediately apparent to the user (Clark and Feng, 2017). The user might then suspect that it is a false alarm, which may lead to the user not responding properly to the TOR.

Based on the meta-analysis by Eriksson and Stanton (2017) and by including recent literature, an analysis is made what time budgets for different take-over situations are used in simulated driving studies (see Table 2.1). The used definition of the type of take-over situation is according to the respective authors of the analysed simulated driving studies. The analysis shows that there is no unambiguous definition of take-over criticality. What does become clear is that the time budget is often defined as the time from the TOR to reaching an obstacle, such as a broken vehicle or construction site. Sometimes, however, the time budget equals the time from the TOR to deactivation of the ADS. For example, Gold et al. (2016) used a 0-second time budget for a critical take-over situation in which the ADS immediately got disengaged when the TOR was issued, even though the Time-To-Collision (TTC) was 7 seconds. Furthermore, in the table can be seen that Mok et al. (2015) and Naujoks et al. (2017) characterise a missing lane markings event with a 0-second time budget differently. Mok et al. (2015) defines it as a critical event, while Naujoks et al. (2017) defines it as a non-critical event. This difference has to with the fact that the event used in the study by Mok et al. (2015) occurs on a curved road section, instead of on a straight road section. According to Naujoks et al. (2017), no immediate action is required by the user if this event occurs on a straight road section, thus it can be defined as a non-critical event. Furthermore, Eriksson and Stanton (2017) used an infinite time budget, this time budget was chosen because their study aimed to analyse how long it took for drivers to take-over the driving task, without any time pressure. Finally, Payre et al. (2016) did not classify the take-over as critical or non-critical, but as anticipated or unforeseen. Alrefaie et al. (2019), Baek et al. (2019), Melcher et al. (2015), Schömig et al. (2015) did not define the take-over type at all.

From the above analysis, not a clear classification of take-over situations can be derived. However, it can be seen that a take-over is generally considered critical when an obstacle or construction site appears in the lane of the ego-vehicle. What can be seen from the table is that the combination of the take-over situation and the duration of the time budget makes a take-over critical or not. As for the time-criticality of the time budget, different definitions are used: it is critical if the TTC is less than 10 seconds or if immediate action is needed by the driver and is defined as uncomfortable if the time budget is less than 8 seconds (Dambock et al., 2013; Melcher et al., 2015; Petermeijer et al., 2016; Naujoks et al., 2017).

Table 2.1: Time budget in highway take-over situations, based on Eriksson and Stanton (2017)

Study	Take-over type	Time budget [s]	Take-over event
Alrefaie et al. (2019)	Not defined	0	No apparent reason
Baek et al. (2019)	Not defined	6*	Obstacle in lane of ego-vehicle
Van den Beukel and Van der Voort (2013)	Critical	0.5*, 1*, 1.5*	Sudden braking lead vehicle
Bueno et al. (2016)	Critical	10*	Obstacle in lane of ego-vehicle
Eriksson and Stanton (2017)	Non-critical	∞	No apparent reason
Feldhütter et al. (2016)	Critical	6*	Obstacle in lane of ego-vehicle
Gold et al. (2013)	Critical	5*, 7*	Obstacle in lane of ego-vehicle
Gold et al. (2016)	Critical	0	Obstacle in lane of ego-vehicle
Kerschbbaum et al. (2015)	Critical	7*	Obstacle in lane of ego-vehicle
Körber et al. (2015)	Critical	3*	Obstacle in lane of ego-vehicle
Lorenz et al. (2014)	Critical	7*	Obstacle in lane of ego-vehicle
Melcher et al. (2015)	Not defined	10*	Appearing construction site
Mok et al. (2015)	Critical	0	Missing lane markings
Naujoks et al. (2014)	Critical	0*	Appearing construction site
Naujoks et al. (2014)	Critical	0*	No apparent reason
Naujoks et al. (2017)	Non-critical	0	Missing lane markings
Naujoks et al. (2018)	Critical	8*	Sudden braking lead vehicle
Payre et al. (2016)	Anticipated	30	Leaving highway
Payre et al. (2016)	Unforeseen	2	ADS failure
Petermeijer et al. (2016)	Non-urgent	>10*	Obstacle in hard shoulder
Petermeijer et al. (2016)	Non-urgent	>10*	Right traffic lane closed
Petermeijer et al. (2016)	Non-urgent	>10*	Appearing construction site
Petermeijer et al. (2016)	Non-urgent	>10*	Traffic jam
Petermeijer et al. (2016)	Non-urgent	>10*	Leaving highway
Radlmayr et al. (2014)	Critical	7*	Obstacle in lane of ego-vehicle
Scharfe et al. (2020)	Non-critical	0	No apparent reason
Schömig et al. (2015)	Not defined	12*	Appearing construction site
Walch et al. (2015)	Critical	4*, 6*	Fog appearing
Yoon and Ji (2019)	Non-emergeny	0	Changing number of lanes
Zeeb et al. (2015)	Critical	4.9*, 5.7*, 6.6*	Obstacle in lane of ego-vehicle
Zeeb et al. (2015)	Non-critical	12*	Appearing construction site
Zeeb et al. (2016)	Non-critical	2.5	Appearing construction site
Zeeb et al. (2016)	Non-critical	4	Missing lane markings

*Defined as the time from issuing the TOR to reaching the obstacle / fog / construction site.
All others: defined as the time from issuing the TOR to deactivation of the ADS.

2.3 WORKLOAD

In the previous section, it was explained that the available time budget plays a critical role in the take-over process. The time budget provided to the user to respond to the take-over request (TOR) must be of sufficient duration in order to enable the driver to respond to the TOR in a timely and safe manner. The available time budget is found to play a role in the experienced workload by the driver, which affects take-over performance (Gold et al., 2013). De Waard (1996) defines mental workload as the amount of workload a driver experiences, which depends on the difficulty of the task. The difficulty is related to the task complexity and to drivers' capabilities, and is thus both task- and driver-specific. Task-specific factors are related to the demand of the processes, e.g. traffic and roadway conditions, whereas driver-specific factors are related to the motivation, strategy, mood and state of the driver.

One of the ironies of automation has to do with low and high workload, which is a concern for safety when driving in automated vehicles (Bainbridge, 1983). Figure 2.3 shows how both little and high demand cause high workload, negatively influencing driving performance. High workload as a result of little demand is known as underload, whereas high workload as a result of high demand is known as overload (Coughlin et al., 2009). De Waard (1996), however, defines underload as a state of high workload, because a larger proportion of the remaining capacity is required for task execution. Mental underload can thus be found in performance region D. In performance region A, the driver can reach high performance of the task execution. In performance regions A1 and A3, however, the driver must make extra effort to maintain high performance. Subsequently, in performance region B, the driver cannot sustain extra effort causing the performance to deteriorate. Eventually leading to a state of mental overload and low performance in performance region C.

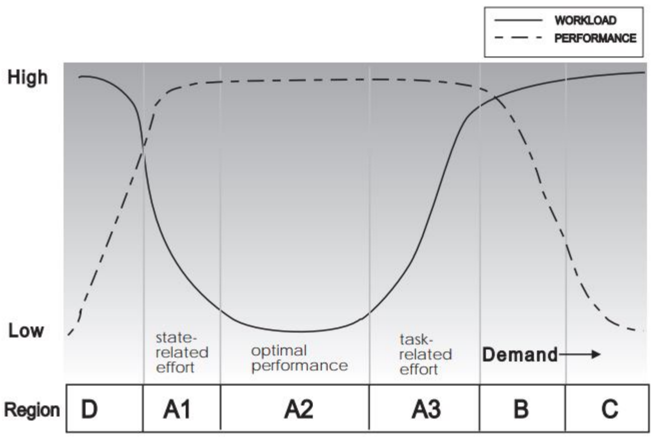


Figure 2.3: Workload and driving performance as a function of task demand (De Waard, 1996)

Mental underload can cause drowsiness, inattention, and slower reactions (Bainbridge, 1983; Endsley, 2019). It can occur when the Automated Driving System (ADS) takes over (part of) the Dynamic Driving Task (DDT), i.e. lateral and longitudinal vehicle motion control and Object and Event Detection and Response (OEDR). As discussed in Section 2.2.1, the user of a vehicle equipped with conditional driving automation, is allowed to engage in Non-Driving Related Tasks (NDRTs) while the ADS is engaged until the user is expected to take over control. Users of such vehicles can, therefore, experience underload when the ADS controls the DDT (Endsley, 2019). The requirement of users of such vehicles to be DDT fallback-ready, is therefore challenging for users if they are in a mental underload state.

As previously discussed in Section 2.2.1, some level 3 vehicles are equipped with a failure mitigation strategy if the user is not responsive to a TOR. Therefore, especially for level 3 vehicles which are not equipped with such a failure mitigation strategy, users' mental underload is an important factor for take-over performance, as underload can cause inability of the driver to execute the DDT safely or to move the vehicle to a minimal risk condition. Studies by De Winter et al. (2014) and Dambock et al. (2013) compared the workload of users in level 3 vehicles, to level 2 vehicles, to level 1 vehicles equipped with Adaptive Cruise

Control (ACC), and to level 0 vehicles without any ADS. Their comparison demonstrated that users of level 3 vehicles are more likely to be involved in an accident in a critical take-over event if they are not attending to the road and prepared to take over. The irony of level 3 vehicles, is therefore the expectation of the user to be the DDT fallback, while being allowed to engage in NDRTs, which can cause the user to experience underload which is detrimental for driving safety.

2.3.1 The role of task engagement

Level 3 vehicles do not require a driver to monitor the ADS (Section 2.1), this therefore allows drivers to engage in NDRTs during automated driving (Gold et al., 2016). However, users of level 3 vehicles still play an important role when driving outside the Operational Design Domain (ODD) of the ADS, as well as when an issue occurs to the ADS, or to the vehicle itself during automated driving (Section 2.2). It was also mentioned that a receptive fallback-ready user has to perform the DDT fallback when requested. However, before being able to perform the DDT fallback, the fallback-ready user has to shift attention to the road and to cognitively process the situation before being able to safely execute the DDT fallback (Petermeijer et al., 2016). These are two necessary steps in the take-over process. The study by Gold et al. (2013) showed that the duration of this last step is critical in terms of safety. Their study showed that receptive users are able to respond quickly to a TOR even with little time for cognitive processing, but the quality of the response is generally lower compared to situations where the user was allowed to have more time for cognitive processing and action selection. Gold et al. (2013), based the quality of the response on the number of mirror gazes and shoulder checks, acceleration and brake usage, where less gazes and checks and increased acceleration and brake usage were qualified as lower quality responses. However, involvement in a NDRT prior to the TOR creates cognitive distraction which slows down these two steps (Radlmayr et al., 2014; Gold et al., 2016). The study by Merat et al. (2014) showed that this is not necessarily an issue when driving in a low complexity driving environment, however drivers take-over performance deteriorates when driving in a complex driving environment (for example, during increased traffic density). When engaged in a NDRT, a TOR can result in a sudden increase in workload which can be detrimental to driving safety (Merat et al., 2014). Clark and Feng (2017) studied the effect of age and NDRT engagement on take-over performance in a situation without any traffic and found that NDRT engagement had no effect on take-over performance, which is in line with the conclusion by Merat et al. (2014).

2.3.2 The role of situation complexity

The traffic density at the take-over location is found to be of great influence on workload and take-over performance (Gold et al., 2016). Their research showed how both low traffic density and high traffic density affect driving performance after the TOR. Deterioration in driving performance was reflected in an increased duration to respond to the TOR, by accelerating and braking more often and excessively and by an increased frequency of lane changes. All in all, both low and high traffic density result in restless driving behaviour after resumption of manual control.

2.3.3 The role of personality

Besides external factors that play a role in workload and take-over performance, such as involvement in a NDRT before the TOR or the take-over location, also intrinsic factors, such as the personality of the driver, could play a role. Namely, people share personality traits that influence how situations are experienced and how people behave accordingly (Fox, 2008). Therefore it can be hypothesised that people with different personality traits experience workload differently.

Multiple models have been created for describing personality based on a set of personality traits. The most common and accepted way of describing personality is by using the five-factor model (FFM) of personality traits, also known as the Big Five. This five-division of core personality traits has been demonstrated by multiple researchers, independently of each other, who all found roughly the same five factors (Fox, 2008). Famous studies are those by Cattell (1957), Goldberg (1990, 1992) and McCrae and Costa (1985, 1987). Table 2.2 gives an overview of the Big Five and provides an explanation of the characteristics for each factor.

Table 2.2: The Big Five personality model (Fox, 2008)

Big Five factors	Characteristics
Extraversion	Tendency to experience positive emotions easily
	Tendency to seek company of others
	Talkative and outgoing
	Energetic
	Tendency to seek stimulation
Agreeableness	Tendency to be compassionate
	Co-operative
	Trusting of others
	Accepting of others
Conscientiousness	Tendency to show self-discipline
	Dutiful and responsible
	Planned rather than spontaneous behaviour
	Aim for achievement
Neuroticism	Tendency to experience negative emotions easily
	Often nervous and anxious
Openness	Appreciation for art and adventure
	Unusual ideas
	Imaginative
	High degree of curiosity

The Big Five represent personality at the highest level (John and Srivastava, 1999). However, it does not imply that personality can be reduced to these five levels only. The levels are composed of multiple lower-order personality traits which describe personality more specifically. In questionnaires used to determine an individual's personality, questions are asked as to whether people identify themselves with those lower-order personality traits. Based on their answer for all those lower-order traits, it can be determined to what extent each Big Five factor is present in the personality of that individual.

The evaluation of the Big Five by John and Srivastava (1999) showed that the model is both replicable and generalisable in language and cultures. Because of these two factors, the Big Five became a common and wide-spread used classification of personality. However, there is considerable criticism of the model. Contradictory to the evaluation by John and Srivastava (1999), there is still criticism of the validity of the model in different cultures. The Big Five is said to be consistent in individualistic cultures, but not in accordance with collectivist cultures (Triandis, 1989; Markus and Kitayama, 1991).

Other critiques are related to the hierarchical order of the Big Five model. Different hierarchical orders are found by DeYoung et al. (2002) and Musek (2007) for example. DeYoung et al. (2002) found correlation between the Big Five factors, such that a shared variance of agreeableness, conscientiousness and neuroticism indicated the existence of a higher-order factor: stability. Shared variance of extraversion and openness also gave an indication of the existence of another higher-order factor of personality: plasticity. These two higher-order factors were called the Big Two. Musek (2007) went one step further and found evidence for a General Factor of Personality (GFP), the Big One. The Big One is defined as the highest-order factor determined by high or low values for the Big Five factors on the one hand, or high or low values for the Big Two on the other hand.

Despite the existence of multiple personality classifications, the Big Five is the most used classification. By using the Big Five as personality classification in this study, use can be made of the Big Five Inventory (BFI), a widely used method for distinguishing people according to their personality. In addition, the results of this study can be compared more easily to other studies that also used the Big Five.

By distinguishing drivers' personalities based on the Big Five, Taubman-Ben-Ari and Yehiel (2012) categorised four driving styles of the Multidimensional Driving Style Inventory (MDSI) that was defined by Taubman-Ben-Ari et al. (2004), namely: the reckless and careless style, the anxious style, the angry and hostile style, and the patient and careful style. It has been found that the reckless and careless style, and the angry and hostile style are both associated with a lower tendency of agreeableness and conscientiousness. People that score high on either agreeableness, conscientiousness, or openness are related more often to the patient and careful driving style. While the anxious driving style is more likely for people that score high on neuroticism and low on conscientiousness.

The attentional control theory by Eysenck et al. (2007) states that people who score high on trait-anxiety (which is part of the neuroticism trait) have an increased allocation of attention to threat-related stimuli compared to other personalities. A TOR can be qualified as threat-related stimulus as it warns the fallback-ready user that the ADS will soon or immediately be stopped. Therefore, it can be hypothesised that a faster response time to a TOR is expected for people who score high on neuroticism, compared to people scoring high on other personality traits.

3 METHODOLOGY

Originally, the aim was to study whether personality plays a role in experienced workload and thereby take-over performance. However, the COVID-19 pandemic made it impossible to carry out a driving simulation experiment with participants, taking into account the measures taken by Delft University of Technology. Therefore, the focus of this study shifted to conducting a validation experiment regarding the set-up of the driving simulator experiment on the role of personality in driver workload induced by a TOR. For this purpose, an $N = 1$ experiment was performed with the researcher as the sole participant. The design of the experiment is validated by varying various factors that are expected to affect workload. Ultimately, this study will provide an empirically validated set of design variables for the study regarding the role of personality traits in automated driving.

This chapter is organised as follows, Section 3.1 provides a short introduction to the experimental approach of the original study regarding the role of personality. After that, the methodology for conducting the $N = 1$ validation experiment is presented in Section 3.2. First the design variables which are varied in the experiment are presented in Section 3.2.1, then in Section 3.2.2 it is explained how driver workload will be measured. The experimental design and schedule are presented in Section 3.2.3 and 3.2.4. After that follows an explanation of the used apparatus in Section 3.2.5. The design of the simulation and the procedure of the simulation are presented in Section 3.2.6 and 3.2.7. After that follows an explanation of the analysis methodology for the various workload measures in Section 3.2.8. At last, the risks and limitations associated with the used methodology are presented in Section 3.2.9.

3.1 STUDY REGARDING THE ROLE OF PERSONALITY

The aim of the original study was to investigate whether personality plays a role in the workload experience of drivers in vehicles equipped with an Automated Driving System (ADS), where the driving task is switched between the ADS and the user. For this, a driving simulator experiment would take place with over 100 participants, differing in personality. This research was approved by the Human Research Ethics Committee of Delft University of Technology (Appendix B).

Participant recruitment

For this experiment, volunteers were recruited through the use of advertisements at the Dutch driving licensing agency (Dutch: CBR), at the Royal Dutch Touring Club (Dutch: ANWB), at Delft University of Technology, and through a personal network. The only requirement to participate was owning a driver's license. There was no requirement for a certain level of driving experience and previous driving simulator experience. Application was open to both Dutch- and English-speaking individuals.

To apply for participating in the driving simulation experiment, volunteers had to fill in an online application form (Appendix C). Because personal data is used in this study, a data management plan was made and approved by the Human Research Ethics Committee of Delft University of Technology. To protect the personal data, the application form was created in Qualtrics. The form contained multiple questionnaires: a demographic questionnaire, a driving experience questionnaire, a health questionnaire, and the 44-item Big Five Inventory (BFI). Forms were available in both Dutch and English. The Dutch version of the BFI is a translation by Denissen et al. (2008) of the original English version by Goldberg (1992) and John (1990) The BFI in both Dutch and English can be found in Appendix C and Appendix D.

Participant selection procedure

The filled-in questionnaires are analysed with MATLAB to determine the personality of the volunteer. It was intended to include the most extreme personalities per trait in the experiment. To do so, the answers to the 44-item BFI were translated into scores for each of the five personality traits. As the BFI uses a five-point Likert scale, answers were translated into scores ranging from 1 to 5 points, where 1 point is awarded to 'strongly disagree' and 5 points to 'strongly agree'. The volunteers' scores for the BFI are calculated using the scoring instructions provided by John et al. (2008). The score for each Big Five factor is calculated by summing the answers to specific questions, see Table 3.1. Some numbers in the table are underlined, meaning that the inverse of the score on that question is used. Thus, 'strongly disagree' is awarded 5 points, whereas 'strongly agree' 1 point. After summing all the scores, the total score for each Big Five factor is normalised by dividing the score by the maximum score that can be obtained for that specific factor. This way a scale score is obtained for every Big Five factor.

Table 3.1: Scoring method 44-item BFI

Big Five factors	**Summation of item numbers BFI**									
Extraversion	1,	$\underline{6}$,	11,	16,	$\underline{21}$,	26,	$\underline{31}$,	36		
Agreeableness	$\underline{2}$,	7,	$\underline{12}$,	17,	22,	$\underline{27}$,	32,	$\underline{37}$,	42	
Conscientiousness	3,	$\underline{8}$,	13,	$\underline{18}$,	$\underline{23}$,	28,	33,	38,	$\underline{43}$	
Neuroticism	4,	$\underline{9}$.	14.	19,	$\underline{24}$,	29,	$\underline{34}$,	39		
Openness	5,	10,	15,	20,	25,	30,	$\underline{35}$,	40,	$\underline{41}$,	44

Every volunteer is awarded a scale scores for each personality trait. Subsequently the volunteer is assigned to their dominant trait (i.e. trait with highest scale score). The 20 highest scoring volunteers per trait are selected to participate in the driving simulation experiment. When participants are selected for the study, they are sent an information sheet and informed consent form. A more detailed explanation of the experiment and procedure is provided, explaining the risks involved in participating, the privacy of the participants, sharing of results, ethical approval and consent form for the previous mentioned details. The participants are asked to read the information sheet and informed consent form carefully. The English and Dutch information sheet and informed consent form can be found in, respectively, Appendix E and F.

3.2 VALIDATION EXPERIMENT

A validation experiment is conducted with the aim to provide an empirically validated set of design variables for the study regarding the role of personality traits in automated driving. In the validation experiment, the driving task is transferred from the Automated Driving System (ADS) to the driver in a Level 3 vehicle under varying task demand conditions. A driving simulation experiment is designed for this purpose, in which the researcher is the only participant in the experiment. To validate the design of the experiment, multiple design variables of the driving scenario will be varied and their effect on take-over request (TOR)-induced workload is analysed. As explained in Section 2.3, workload depends on the difficulty of the task, which in turn depends on two factors: task complexity and driver competence. If in this validation experiment it can be validated that the measured workload reflects the degree of task complexity with only one driver, variations in workload in an experiment regarding the role of personality in workload and take-over performance can be attributed to changes in driver competence, in which personality may play a role.

3.2.1 Design variables varied in the experiment

The task complexity of the driving simulation experiment depends on several design attributes that affect demand processes of the scenario. To measure the extent to which workload is experienced at a TOR, the design of the take-over is of great importance. The design attributes that are chosen to be varied in the experiment are the time budget of the TOR, the traffic density and road shape at the TOR location, and whether the user engages in a Non-Driving Related Task (NDRT) during automated driving. An overview of the design attributes and their attribute levels varied in the experiment can be found in Table 3.2. It is chosen to vary the four design attributes because these attributes are expected to lead to measurable differences in task demand and complexity, and, in turn, lead to differences in workload measurements.

Table 3.2: Demand processes to be varied in the experiment

Design attributes	Attribute levels
Time budget of the TOR	0 seconds 5 seconds 10 seconds 15 seconds
Traffic density at TOR location	0 vehicles/km/lane 5 vehicles/km/lane 10 vehicles/km/lane 15 vehicles/km/lane
Location of the TOR	Straight road Curve
Task during automated driving	Monitoring Playing a game

As elaborated in Section 2.2.2, take-over behaviour and performance is related to the time budget that is provided to the user to respond to the TOR. If the time budget is too limited, a lower quality take-over performance is expected (Gold et al., 2013). By varying the time budget, different degrees of task complexity can be simulated. Four attribute levels are chosen, which are 0, 5, 10 and 15 seconds. With these attribute levels, the effect of both urgent as non-urgent time budgets on workload and take-over performance can be investigated. Here, urgent is defined as a time budget of less than 10 seconds and non-urgent as 10 seconds or greater, as defined by Petermeijer et al. (2016). Moreover, the 0 and 15-second time budget are included as attribute level, as it provides insight into the effect on workload of, respectively, no time budget or a very great time budget. It is chosen to use a 15-second time budget instead of an infinite time budget as Eriksson and Stanton (2017) did, in order to assure orthogonality of the attribute (more information on orthogonal designs is provided in Section 3.2.3). As an average take-over reaction time (TOrt) of 2.06 and 3.10 seconds for 5- and 7-second time budgets was found by Gold et al. (2013), it is expected that a 15-second time budget is sufficient for drivers to take-over the driving task.

The traffic density at the take-over location is found to also be of influence on workload and take-over performance (Gold et al., 2016). For this design attribute, the following equidistant attribute levels are chosen: 0, 5, 10 and 15 vehicles per kilometre, which are defined as a, respectively, zero, low, medium and high traffic density. A traffic density of 15 vehicles per kilometre is used, instead of greater traffic densities, as in the study by Stanton and Young (2005) a ceiling effect above this traffic density on workload was found. Moreover, traffic flow becomes unstable with traffic densities greater than 18 vehicles per kilometre at a speed of 120 km/h (Schöpplein, 2013, as cited in Gold et al., 2016). By including zero, low, medium and high traffic densities as attribute levels, it can be studies whether increasing traffic densities affect workload to such an extent that take-over performance deteriorates, as can be expected from literature.

The third attribute that is varied in the experiment is the road shape at the take-over location. As mentioned previously in Section 2.2.2, a take-over can be defined as non-critical if it occurs on a straight road section, and critical if it occurs on a curved road section (Naujoks et al., 2017; Mok et al., 2015). As the driver has to take immediate action when a TOR is issued in a curve, this take-over is expected to have a greater effect on workload and thereby lower quality take-over behaviour can be expected.

The fourth and final attribute that is varied in the experiment is whether the driver is engaged in a NDRT during automated driving. According to the requirements set by SAE International (2018), users of level 3 vehicles are allowed to engage in NDRTs during automated driving. As elaborated in Figure 3.2.8, various NDRTs are used to study the effect on, amongst others, take-over performance. It is chosen to use Tetris as NDRT, as it assures engagement in the task by requiring continuous visual and cognitive attention.

3.2.2 Measuring driver workload

To measure driver workload, different types of measures can be used that differ in their sensitivity to workload at different levels of task demand (De Waard, 1996; Mehler et al., 2011) Four types of workload measures can be distinguished: subjective and physiological workload measures, driving performance indicators, and secondary task performance measures (De Waard, 1996). In order to gain a complete image of workload at different levels of task demand, a set of four workload measures is used in the experiment.

Subjective workload

Commonly used subjective workload measures are the Subjective Workload Assessment Technique (SWAT) and the NASA Task Load Index (TLX) (Rubio et al., 2004). In this study it is chosen to measure subjective workload by means of the Raw Task Load Index (RTLX), which is a simplification of the NASA TLX. The RTLX measures workload on six scales: mental demand, physical demand, temporal demand, performance, effort, and frustration. The TLX uses the same scales, but requires pairwise comparisons between all the scales, making it a more time-consuming measure than the RTLX, which only calculates the mean value of the six scales (Hart and Staveland, 1988; Hart, 2006).

Physiological workload

Popular physiological measures for workload are heart activity (Heart Rate (HR) and Heart Rate Variability (HRV), blink rate, breathing rate, Skin Conductance Level (SCL) (also known as Galvanic Skin Response (GSR)) and brain activity (Van Gent et al., 2017; Mehler et al., 2011). HR is a commonly used measure in studies examining workload in response to changes in driving demand. As measuring heart activity is low-cost and can be done in non-intrusive way with easy-to-use measures, it is chosen to analyse heart activity as physiological workload measure in this study (Van Gent et al., 2017). Lohani et al. (2019) reviewed studies that researched HR and HRV as a function of workload. They concluded that HR increases with workload as a result of increasing cognitive demand, and HRV decreases with workload in increasing task demands. Furthermore, it has been found that "HRV is sensitive to workload increases due to vigilance and situational awareness demands of the task". Some studies, however, found no such relationship between workload and HR and HRV, e.g. Shakouri et al. (2018), who found that only subjective workload measures showed differences in workload due to task demands. Despite this result, Lohani et al. (2019) conclude that both HR and HRV can be used as workload measure in changing task demand settings. They urge nonetheless, that it is important to consider contextual factors that can impact the measurement of HR and HRV. For instance, it has been found that both heart activity measures have a time-on-task effect. As the task becomes less difficult over time, the participants become more relaxed, disengaged or demotivated, influencing the measures. Moreover, the study by Paxion et al. (2014) found that HR also reflects energetic, thermoregulatory, respiratory, emotional processes, as well as emotional strain and physical activity.

In literature, there are many different HRV parameters used to measure changes in workload. These measures use the R-to-R Interval (RRI), which is derived from the QRS-complex of the heartbeat, see Figure 3.1. The RRI is also known as Peak-to-Peak Interval (PPI), Inter-Beat Interval (IBI), or Normal-to-Normal Interval (NNI) (Selvaraj et al., 2008; Paxion et al., 2014). From the RRI, the Low-Frequency (LF)/High-Frequency (HF) ratio can be derived, which provides a robust overall method to assess workload (Mehler et al., 2012). HRV measures can be categorised in two groups: time-domain and frequency-domain measures. Common time-domain measures are Standard Deviation of Normal to Normal intervals (SDNN), Root Mean Square of Successive Differences (RMSSD) and NN50 which is defined as the number of adjecent NNIs that differ more than 50ms. With regards to frequency-domain measures, common measures are the Ultra-Low-Frequency (ULF), LF-, and HF-domain, and the LF/HF-ratio (Task Force of the European Society of Cardiology and Electrophysiology, 1996).

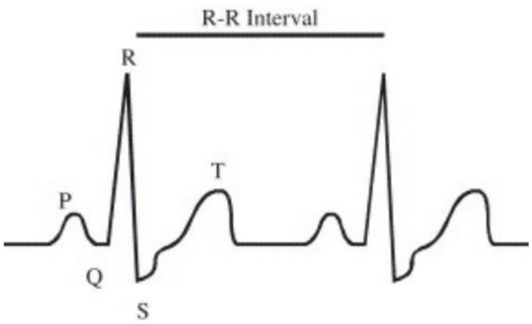


Figure 3.1: The QRS-complex of the Electrocardiogram (ECG) (Peterkova and Stremy, 2015)

For this study it is chosen to include the following measures: HR in beats per minute (bpm), HRV indicated by RMSSD, SDNN and HF-power. This way, both frequency and time-domain HRV measures will be included. Mean HR and changes in HR after a TOR are a commonly used workload measure (Alrefaie et al., 2019; Mehler et al., 2012). However, using the RMSSD and SDNN as workload measure is a novel approach for measuring TOR-induced workload. Currently there is no consensus on the usefulness of these HRV measures as workload measure Mehler et al. (2011); Luque-Casado et al. (2016); Hidalgo-Muñoz et al. (2019). As far as is known, only Pakdamanian et al. (2020) (published at the time of writing this study) used RMSSD and SDNN to measure TOR-induced workload. Pakdamanian et al. (2020) conducted an exploratory study with two participants who experienced four TORs under two weather conditions (sunny / rain) and alert modalities (visual-auditory / auditory). Other studies used the RMSSD and/or SDNN to measure workload differences between low and high task demands during automated driving Mehler et al. (2011); Luque-Casado et al. (2016); Hidalgo-Muñoz et al. (2019); Heine et al. (2017); Shakouri et al. (2018); Heikoop et al. (2018). There is no consensus yet on using these indicators as workload measure, as these measures did not always indicated the expected direction of the effect on workload. HF-power will be used as substitute to the LF/HF-ratio, which is one of the most common HRV indices for measuring TOR-induced workload. However, as the LF/HF-ratio requires at least 5 minutes of data, it is not suitable to use in a driving simulation experiment with a short duration of driving before and after the TOR (Section 3.2.6) (Shaffer and Ginsberg, 2017). The HF requires only 1 minute of data, which is therefore suitable to use for simulations with a short duration. Moreover, the HF was found to measure workload reliably in high task demand situations (Mehler et al., 2011). As the TOR is expected to result in a high task demand, it is expected that the HF will prove to be a valuable workload measure.

Driving performance

Driving performance indicators will be used, which provide insight into the driver reaction-time to the TOR and if and when driving performance is affected during manual driving after the take-over. Commonly used indicators in studies regarding transitions of control can be grouped into three groups. Firstly, longitudinal control ability indicated by the minimum, maximum, mean and Standard Deviation (SD) speed. Secondly, lateral control ability indicated by the Standard Deviation of Lateral Position (SDLP) and steering wheel movements (i.e. maximum angle, maximum angle, number of 1° reversals, number of corrections). Thirdly and lastly, time indicated by the TOrt and Time-To-Collision (TTC) (Reimer et al., 2012; Merat et al., 2014; Bueno et al., 2016; Varotto, 2018). For this study, it is chosen to use a combination of these indicators: the SD speed, the SDLP and the TOrt. The SD speed is used, as it is one of the most common driving performance indicators in workload studies. Furthermore, the SDLP is used as the steering wheel of the used driving simulator is not able to sufficiently accurately log movements to the steering wheel, whereas SDLP requires a less accurate calculation method based on the position of the ego-vehicle on the XYZ-plane logged by the simulator. The TOrt is used as in this study non-critical transitions of control will be simulated.

Secondary task

By using a secondary task during automated driving, cognitive distraction from the driving task can be simulated. For instance, with the commonly used *n*-back task a series of numbers are verbally presented to the driver and the driver has to recall from memory and respond with the n'th number that was presented before the current number (Mehler et al., 2009). However, for this experiment, playing the game Tetris is used as secondary task. The advantage of this task over other tasks, such as the *n*-back, is that it is possible to engage in this task without the requirement of any other researcher to prepare the task or be involved during the experiment; so it is suitable for a self-experiment. In addition, while playing Tetris, the user cannot be distracted from the game and still monitor the road, so it ensures that the user is continuously involved in the task (Section 3.2.1).

In sum, the following measures are used:

1. Subjective workload: RTLX, a simplification of the NASA TLX;
2. Physiological workload: HR in bpm and RMSSD, SDNN and HF-power as HRV indicators;
3. Driving performance indicators: TOrt and longitudinal (SD speed) and lateral control ability (SDLP);
4. Secondary task performance: Tetris score.

3.2.3 Experimental design

A full factorial design of the simulation runs is used to vary the four design attributes in the experiment, A full factorial design ensures uncorrelated attributes and preserves attribute level balance. Furthermore, it prevents multicollinearity as it leads to zero correlations between attributes, which results in low standard errors. Because of attribute level balance, all attributes have the same standard error, and thus have the same level of precision. A full factorial, orthogonal, experimental design is created using Ngene (ChoiceMetrics, 2018). Ngene is a software for designing experimental designs, mostly used for designing stated choice experiments. It allows designing profiles in which attributes and their respective levels are varied between choice alternatives, or in this case between simulation runs.

Ngene software provides many different design options, from designing full or fractional factorial designs, to orthogonal or efficient designs. For this validation experiment, it is chosen to make use of a full factorial, and thus orthogonal, design. An advantage of this design type over other design types is that it allows to estimate both main and interaction effects. A fractional factorial design does not allow to estimate interaction effects. A downside is, however, that using a full factorial design results in more profiles than using a fractional factorial design, thus it takes more time to complete the validation experiment. The full factorial design generated by Ngene consists of 64 profiles: multiplying the attribute levels of the chosen design attributes of Table 3.2 gives $4 * 4 * 2 * 2 = 64$ profiles. An orthogonal fractional factorial design would have yielded 48 profiles for the chosen design attributes (ChoiceMetrics, 2018). However, as including interaction effects is necessary for this validation experiment it is chosen to use a full factorial design instead of an orthogonal fractional factorial design. By estimating interaction effects, it is possible to determine whether certain combinations of design attributes and levels lead to an even greater increase in the workload experience. Thus, with a full factorial design, the effect of increasing task complexity on workload can be measured.

The experiment consists of 64 scenarios in which a single TOR is simulated. The experimental design of the 64 scenarios is provided in Table 3.3. In the schedule, the scenarios are listed in a randomised order to prevent bias because of order effect. Only one take-over is included per run, as this minimises the time-on-task effect that biases the workload measurements. A pause between the runs in which the data from the run will be exported and the next run is set up will minimise the time-on-task effect. Moreover, if more take-overs per run would be included, the simulation had to be paused for completing the RTLX or it had to be completed during automated driving after the hand-over. Completing the questionnaire during automated driving limits the duration of monitoring or playing Tetris. Now, the RTLX will be completed after finishing the simulation run. Therefore there will be no interference with the NDRT, as the time required for completing the questionnaire varies and could thus bias the workload measurements.

Table 3.3: Experimental design of the validation experiment

	Profile scenario					Profile scenario			
#	TB*	TD*	Location	Task	#	TB*	TD*	Location	Task
1	10	5	Straight	Monitoring	**33**	15	10	Curve	Tetris
2	0	0	Curve	Tetris	**34**	5	10	Curve	Tetris
3	5	5	Straight	Monitoring	**35**	10	0	Straight	Monitoring
4	15	10	Straight	Monitoring	**36**	15	10	Curve	Monitoring
5	15	15	Curve	Tetris	**37**	15	0	Curve	Monitoring
6	5	5	Curve	Tetris	**38**	10	15	Curve	Tetris
7	10	5	Curve	Monitoring	**39**	5	15	Straight	Tetris
8	5	10	Straight	Monitoring	**40**	5	5	Curve	Monitoring
9	5	5	Straight	Tetris	**41**	0	15	Straight	Monitoring
10	5	15	Curve	Tetris	**42**	10	10	Straight	Monitoring
11	15	15	Straight	Tetris	**43**	0	5	Curve	Monitoring
12	15	15	Curve	Monitoring	**44**	15	10	Straight	Tetris
13	10	15	Curve	Monitoring	**45**	10	10	Curve	Monitoring
14	0	5	Straight	Tetris	**46**	10	5	Straight	Tetris
15	0	10	Curve	Monitoring	**47**	0	10	Straight	Monitoring
16	0	15	Straight	Tetris	**48**	0	0	Curve	Monitoring
17	15	15	Straight	Monitoring	**49**	5	10	Curve	Monitoring
18	5	0	Straight	Monitoring	**50**	0	10	Curve	Tetris
19	0	0	Straight	Monitoring	**51**	10	10	Straight	Tetris
20	0	15	Curve	Monitoring	**52**	5	0	Curve	Monitoring
21	15	0	Straight	Monitoring	**53**	10	0	Curve	Tetris
22	0	0	Straight	Tetris	**54**	5	0	Straight	Tetris
23	10	0	Curve	Monitoring	**55**	15	5	Straight	Tetris
24	15	5	Curve	Monitoring	**56**	0	15	Curve	Tetris
25	15	0	Straight	Tetris	**57**	15	5	Curve	Tetris
26	10	0	Straight	Tetris	**58**	0	10	Straight	Tetris
27	0	5	Straight	Monitoring	**59**	15	0	Curve	Tetris
28	5	15	Curve	Monitoring	**60**	10	15	Straight	Tetris
29	10	15	Straight	Monitoring	**61**	5	15	Straight	Monitoring
30	15	5	Straight	Monitoring	**62**	5	0	Curve	Tetris
31	10	5	Curve	Tetris	**63**	10	10	Curve	Tetris
32	5	10	Straight	Tetris	**64**	0	5	Curve	Tetris

* TB = Time budget [s], TD = Traffic density [vehicles/km/lane]

3.2.4 Experiment schedule

As experimenting days were limited due to the COVID-19 circumstances, a daily schedule was made (Table 3.4). For every experiment run, 20 minutes were scheduled for setting up of the experiment, conducting the experiment, and finalising (i.e. data exporting). In the schedule, experimenting was scheduled on five alternating days over two weeks which were limited from 9AM to 5PM because of the campus opening hours of Delft University of Technology. In the schedule, the first 30 minutes of the day are scheduled for setting up the first runs, which takes longer than setting up other runs because the driving simulator must be started and the equipment needs to be set up for the day. It was expected that three runs could be completed every hour. After every hour a 15-minute break was scheduled, as well as a 45-minute lunch break after the first six runs of the day. The last few hours of the day, a 15-minute break was scheduled after every two runs instead, to account for fatigue that occurs after performing many simulation runs. At the end of the day, time was scheduled for turning off the driving simulator and for disinfecting all equipment. All things considered, it was expected that 16 runs could be completed per day. Therefore, the experiment could have been completed in four days, if everything went well. The final day of experimenting was scheduled as slack time in case some issues occurred during experimenting or with the data collection.

Table 3.4: Daily schedule validation experiment, example day 1: April 29, 2020

Duration		Task
09:00	09:30	*Start of day: preparing driving simulator and equipment*
09:30	10:30	Run 1-2-3
09:30	*09:35*	*Set up run 1*
09:35	*09:45*	*Driving run 1*
09:45	*09:50*	*Completing run 1: NASA RTLX & data export*
09:50	*10:10*	*Run 2*
10:10	*10:30*	*Run 3*
10:30	10:45	*Break*
10:45	11:45	Run 4-5-6
11:45	12:00	*Break*
12:00	13:00	Run 7-8-9
13:00	13:45	*Lunch break*
13:45	14:45	Run 10-11-12
14:45	15:00	*Break*
15:00	15:40	Run 13-14
15:40	15:55	*Break*
15:55	16:35	Run 15-16
16:35	17:00	*End of the day: turn off driving simulator and disinfect equipment*

Next experiment days: May 1, 6, 8 & 12, 2020

3.2.5 Apparatus

The driving simulator that is used for this study is located at the department of Transport and Planning at Delft University of Technology. The simulator is fitted with three high resolution screens providing a 180° field of view. As depicted in Figure 3.2, the three screens simulate the two side windows and the windshield of the vehicle. Furthermore, the simulator is equipped with a Fanatec haptic steering wheel and clutch, brake and gas pedal, a gear stick, hand break, car seat and seat belt. Unfortunately, the vibration function in the driver's seat does not work. This reduces the sense of reality of the driving simulation. Thus, no vibrotactile feedback can be used for the TOR, which was expected to reduce the TOrt (Petermeijer et al., 2016; Eriksson and Stanton, 2017). Simulator data is logged at 50Hz.

A tablet was placed on a holder on the right side of the driver's seat, used for playing Tetris as NDRT during automated driving.

Figure 3.2: Driving simulator used for the experiment

For measuring physiological workload, an optical sensor is mounted on the participant's right index finger that measures light transmission through the fingertip, as can be seen in Figure 3.3 (Allen, 2007). With this optical sensor a Photoplethysmography (PPG) can be obtained, which is logged at 100Hz. An Atmel ATMega328p embedded processor board powers the recording of the data.

Figure 3.3: Finger sensor used for the experiment

Traditionally, the ECG is the standard for measuring heart activity (Selvaraj et al., 2008). However, it is chosen to use PPG for this study as it is a low-cost measure and it measures heart activity in a less intrusive manner as it only required placing a sensor on the participants' fingertip instead of placing multiple electrodes on the participants' chest. This makes PPG also less less time-consuming compared to ECG. A disadvantage of PPG over ECG is, however, that it is sensitive to recording ambient noise van Gent et al. (2019). By using HeartPy, an algorithm to handle heart activity data from PPG developed by van Gent et al. (2019), the data can be filtered and estimates for HR and HRV can be obtained, making PPG a valid alternative for ECG. For noise filtering, a low-pass Butterworth filter with a cutoff of 3Hz is used to remove all data points that resulted in a HR of 180 bpm and greater.

Although HeartPy can filter noisy PPG data, large movements to the sensor may result in data collection problems which HeartPy cannot solve. For this, the Polar H10 chest strap is used as an additional method to measure HR (Figure 3.4). Polar H10 data is logged at 1Hz, which allows chest strap data to only be used as substitute for mean HR data when data collection issues have occurred with the finger sensor. Chest strap data is used when the finger sensor data deviates more than 3 bpm (after the data filtering) from the mean HR measured by the chest strap Jo et al. (2016).

Figure 3.4: Polar H10 chest strap used in the experiment Polar (2020)

The RTLX is used for measuring subjective workload, which is a simplification of the NASA TLX. The RTLX measures workload on six scales: mental demand, physical demand, temporal demand, performance, effort, and frustration. The scale ranges from low (0%) to high (100%), except for the performance scale that ranges from good (0%) to poor (100%). The TLX uses the same scales, but requires pairwise comparisons between all the scales, making it a more time-consuming measure than the RTLX, which only calculates the mean value of the six scales (Hart and Staveland, 1988; Hart, 2006). The scales range from low (0%) to high (100%), except performance that ranges from good (0%) to poor (100%) and are defined by Hart and Staveland (1988) as follows:

- Mental demand: ”How much mental and perceptual activity was required (for example, thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, forgiving or exacting?”
- Physical demand: ”How much physical activity was required (for example, pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?”
- Temporal demand: ”How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?”
- Performance: ”How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?”
- Effort: ”How hard did you have to work (mentally and physically) to accomplish your level of performance?”
- Frustration level: ”How insecure, discourages, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?”

The RTLX is filled in on the tablet by the participant after completion of the simulation run. The ”NASA-TLX” iPhone application is used (NASA, 2016), see Figure 3.5 for an impression.

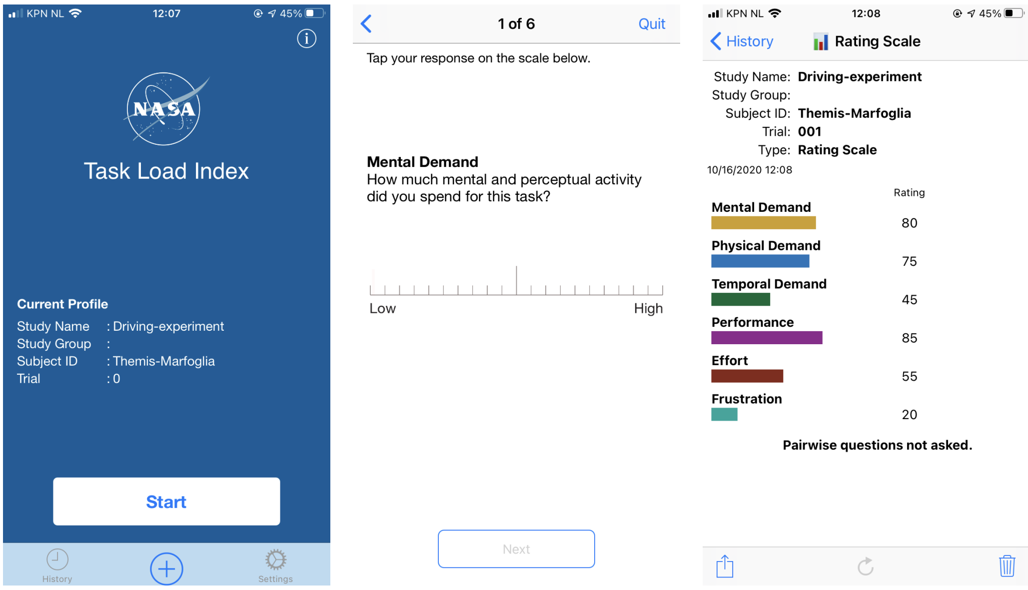


Figure 3.5: Impression "NASA-TLX" iPhone application (NASA, 2016)

3.2.6 Design of the simulation

The driving simulation scenario is developed with Unity game-development software, using C# programming language. Relevant scripts that were developed for this driving scenario are available via 4TU, DOI: 10.4121/13102763.

Road and road environment

For the experiment, the fictive Dutch A3 highway is simulated, which is designed as a 2x2 lane highway. The highway has multiple on- and off-ramps to the city of Amsterdam and fictive towns, named Achthof and Brugstad. Halfway along the highway, the highway has a large loop, after which it crosses the highway again via a viaduct. This loop is an essential element in the design of the experiment, as this allows to validate the combination of a 15-second time budget with a TOR that is issued in a curve, without the vehicle not being located in the curve after the simulated ADS disengages after 15 seconds. As experiment days were limited, it was chosen to design scenarios with a short duration. In total, the highway measures 7 kilometre for scenarios with a TOR on a straight road section, and measures 5.5 kilometre for scenarios with a TOR in a curve. The distance is measured from the start to the end point of the experiment on the highway, see Figure 3.6.
Trees and buildings are located near the highway in order to simulate a realistic highway environment. Moreover, high-mast lighting and guard rails are placed on the sides of the highway and in the central reservation between the lanes. The lanes of the highway have a width of 3.50 meters, which is in accordance with the Dutch Guidelines Design Motorways (Dutch: ROA) (Rijkswaterstaat, 2019). An impression of the scenario can be found in Figure 3.7.

Ego-vehicle and other traffic

During the experiment, four different traffic densities are simulated: 0, 5, 10, and 15 vehicles per kilometre per lane. These traffic densities are only applied to the two lanes in the direction of the ego-vehicle. No traffic is driving in the opposing direction in order to reduce CPU usage. Moreover, behaviour of traffic was limited to only being able to keep to the set following distance. Vehicles, except the ego-vehicle, were not able to overtake other slower driving vehicles, as this caused the simulation to slow down. Thus, it was not possible to simulate vehicles varying in speed. To nevertheless design a more realistic

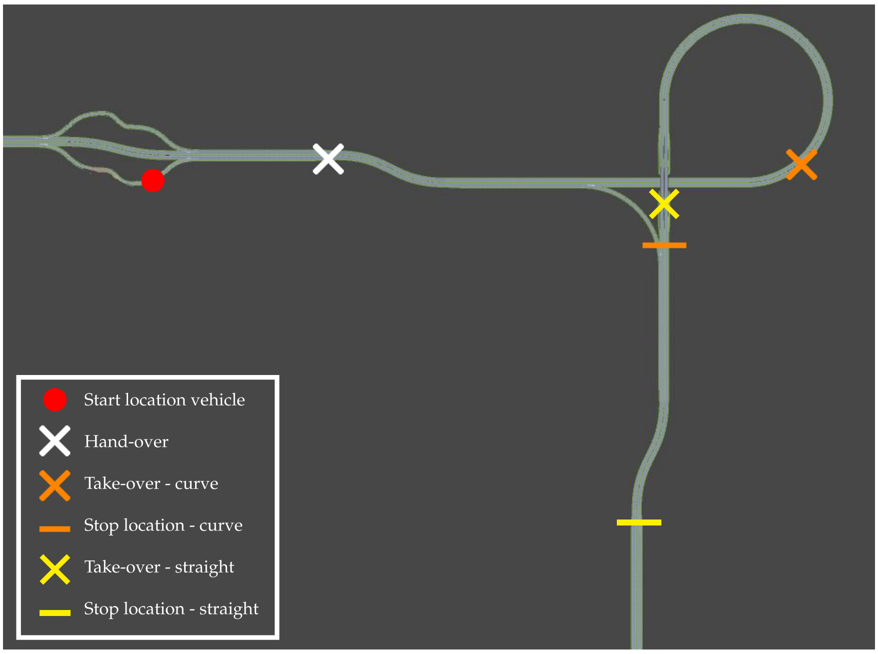


Figure 3.6: Highway layout of the driving simulation experiment

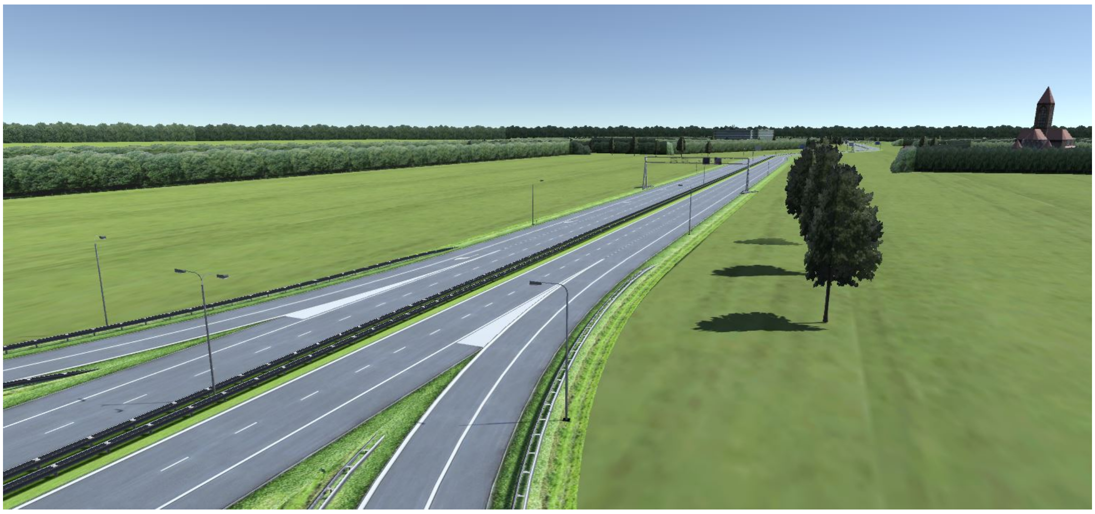


Figure 3.7: View on the highway

driving environment with vehicles overtaking other slower vehicles, it was chosen to set a speed limit of 100 km/h for vehicles on the right lane and 120 km/h on the left lane. Therefore, vehicles were either spawned on the left or right lane, kept to that lane and drove at the set speed limit. Traffic is arranged with a constant and equal distance between the vehicles of 200, 100 and 66.6 meters for a traffic density of, respectively, 5, 10 or 15 vehicles per kilometre. To reach this constant and equal distance, the vehicles were spawned at a regular interval. For low traffic density scenarios, the spawn delay was set to 6 seconds for vehicles on the left lane and 7 seconds for vehicles on the right lane. The spawn delay for moderate traffic density scenarios is set to 3 seconds for the left lane and 3.5 seconds for the right lane. For the high traffic density scenario, the spawn delay was set at 1.5 and 2 seconds for respectively, the left and right lane.

The participant drove a Ford Focus ST with automatic transmission. At the start of the experiment, the vehicle is located on the on-ramp, as illustrated in Figure 3.6. The dashboard is equipped with a speedometer and tachometer. The car is equipped with the regular mirrors, i.e. interior mirror and left and right exterior mirrors. The vehicle is not equipped with a navigation system as this is not necessary for the experiment. During automated driving, the ego-vehicle adheres to the maximum speed of 120 km/h if possible. In the case of slower traffic driving in front of the ego vehicle, the simulated ADS adjusts its speed to maintain a following distance of at least 1 second (at a speed of 120 km/h this equals 50 meters). If possible, the vehicle will overtake slow-moving traffic.

When the ego-vehicle during manual driving drives between all other traffic, the vehicles were set to always maintain the predefined following distance. During overtaking, when the ego-vehicle is moved from the right to the left lane, traffic driving behind the ego-vehicle will decrease their speed temporarily to reach the set following distance. Therefore, rear-end collisions caused by other traffic are not possible in the current design of the simulation. So, even with a long TOrt, no traffic will collide with the ego-vehicle.

Hand-over and take-over

During the experiment, the participant experiences one hand-over from manual to automated driving and one take-over from automated to manual driving. The script used for simulating the hand-over and take-over are available via 4TU, DOI: 10.4121/13102763. The triggers are set at predefined locations in the scenario, as illustrated in Figure 3.6. It was decided to place the take-over on the straight and curved road section close to each other, in order to prevent the driver from predicting whether the take-over would take place in a straight or curved section, in order to prevent bias. Moreover, the design of this validation study created a requirement of locating the TOR of the curved road section in a location where after the 15-second time budget the vehicle would still be located on the curved road. In addition, the take-over that is issued on a straight road section is located after the bridge, in order to prevent additional task demand by issuing the TOR on the bridge which reduces the view on the road. However, an entry road is located near this TOR location, but it is expected that this would not lead to additional task demand as no traffic is driving on this entry road.

During the experiment runs, a take-over is simulated, where the fallback-ready user (i.e. participant, in this case the researcher) is expected to take over the driving task. The simulated take-over is not triggered by an obstacle on the current lane which the ADS cannot handle, such as road works or a stationary car (see table Table 2.1), but is triggered with no apparent reason as is done before in, for example, the studies by Eriksson and Stanton (2017) and Naujoks et al. (2014). The advantage of such a non-critical take-over, is that it allows to vary the time budget provided to the user in the experiment. The driver is notified of the TOR by an auditory signal after 2:00 minutes or 2:30 minutes of automated driving when issued in a curve or straight road section, respectively. The used duration of automated driving before the TOR, as well as the duration of manual driving after the take-over is based on previous driver simulation studies regarding take-overs in a highway setting (Van den Beukel and Van der Voort, 2013; Bueno et al., 2016; Gold et al., 2016; Körber et al., 2016). In this study, the duration of automated driving before the take-over equals 03:30 minutes when the TOR was issued on a straight road section and equals 02:45 minutes when the TOR is issued on a curved road section. The duration of manual driving after the take-over equals 1 minute, which is the same duration as used in the studies by Bueno et al. (2016), Gold et al. (2016) and Körber et al. (2016). As in this study the effect of a take-over on driver workload is measured by means of driving performance measures, among other things, a period of manual driving is necessary after the take-over.

3.2.7 Simulation procedure

Before starting the experiment, the finger sensor and chest strap were set and checked for correct placement. At the start of the experiment, the vehicle was located on the on-ramp. The participant was asked to merge onto the highway when feeling ready. During manual driving, the participant is allowed to choose their own preferred speed, even though they are asked to adhere to the speed limit of 120 km/hand follow regular Dutch traffic laws. As the vehicle is not limited to 120 km/h, it is possible to drive below or above the speed limit. After approximately 30 seconds of manual driving, a hand-over took place which shifts manual to automated driving. The driver would then proceed by monitoring the simulated ADS and vehicle or plays Tetris until the TOR. Without an apparent reason to the driver, the driver was notified of the TOR by an auditory signal after 2:00 minutes or 2:30 minutes when issued in a curve or straight road section, respectively. The driver was asked to take-over as soon as safely possible. To take-over before expiration of the time budget, a key-combination had to be pressed on the keyboard, which was located on the dashboard of the driving simulator. The duration of manual driving after the take-over equalled 1 minute, which is the same duration as used in the studies by Gold et al. (2016); Bueno et al. (2016) and Körber et al. (2016). After the run, the RTLX was filled in by the participant to assess subjective workload.

3.2.8 Driver workload analysis

Table 3.5 provides an overview of the variables that have been collected during the driving simulation experiment. Using these variables, driver workload is assessed by the various workload measures. Here follows an overview of the workload measures and calculation methods.

Table 3.5: Data and variables collected in the validation experiment

Experiment data	Source	Collected variables
Subjective workload	RTLX	Mental demand
		Physical demand
		Temporal demand
		Performance
		Effort
		Frustration
Physiological workload	Chest strap	HR
	Finger sensor	HR
		HRV
Performance measures	Driving simulator	Position ego-vehicle
		Lane number
		Driving mode [automated/manual]
		Position surrounding traffic
Secondary task	Tetris	Tetris score

Subjective workload

For subjective workload, the filled in RTLX questionnaires are used:

- Overall workload score: obtained by averaging the scores on the six scales of the RTLX (Hart, 2006). Scale ranges from 0% to 100%, where 100% is the maximum subjective workload.

- The scores on the six scales will also be presented for the various design variables. Again, the sale range from 0% to 100%, where 100% is the maximum subjective workload.

Physiological workload

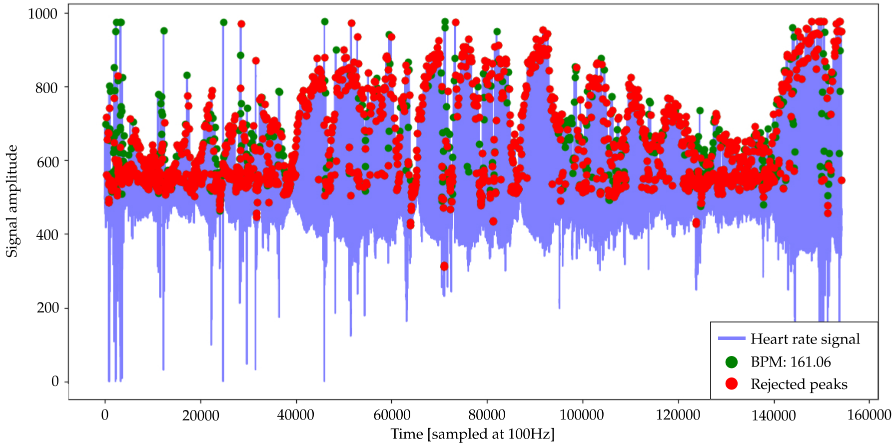
The data that has been collected during the experiment cannot always be used directly to calculate the workload metrics. Therefore, the raw data must first be translated into workable data. In particular, the collected data by the finger sensor must be validated, because the finger sensor is sensitive to ambient noise from bumps to the sensor during the take-over manoeuvre. It is even possible that as a result of bumping the sensor, the sensor does not longer record data. Moreover, as the finger sensor is sensitive to the blood pressure in the tip of the finger, it can occur that the sensor is not able to record when the blood vessels are too small. Therefore, checking the finger sensor data is necessary. First of all, a visual check

of the data is done. Figure 3.8a, provides an example of the worst PPG HR data that has been collected during the experiment. The figure shows many rejected peaks that affect the calculation of the HR and HRV workload metrics. See also Figure 3.8b, which illustrates that most peaks are rejected by HeartPy.
HeartPy rejects a peak if it results in a significantly smaller or bigger RRI than the average. When many peaks are rejected, this can lead to deviating HR values calculated by HeartPy from the true values. In the case of run 7 (see Figure 3.8a and Figure 3.8b) it turned out that much ambient noise had been recorded. As noise generally has a different frequency than the heartbeat, it can be filtered out of the data. To do so, a low-pass Butterworth filter with a cutoff of 3Hz is used to remove all data points that result in a HR of 180 bpm and greater. If after the data filtering, the finger sensor data deviated more than 3 bpm from the chest strap data, the chest strap data is used as substitute for the mean HR (Jo et al., 2016). It cannot be used as substitute for the effect of the TOR on HR, nor for the mean or effect of the TOR on the various HRV indices. Figure 3.8 illustrates the effect of applying the Butterworth filter on noisy data.

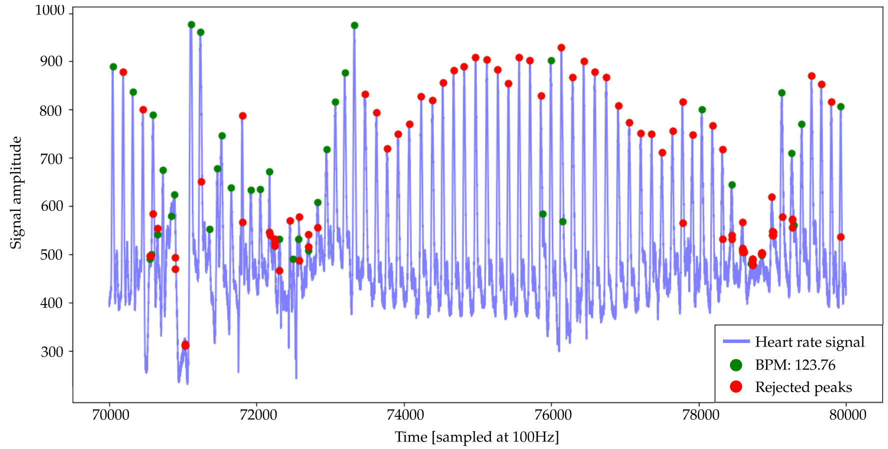
A moving window approach is often chosen as approach to analyse HR and HRV measurements. In order to analyse a certain duration of data, the data is cut in multiple sections that are averaged. For instance, 5 minutes of HR data that is cut in 5 1-minute sections for which the average SDNN is calculated. The 5 1-minute SDNN values are then averaged to get the average SDNN value of the 5 minute section. As a result of using this approach, the curve of found data points is smoothed, which mitigates the effect of outliers. For this study, however, it is not suitable to use a moving window approach, because the duration of the experiment is so limited that a single window will yield more reliable data than averaging multiple smaller windows (Munoz et al., 2015; Cho et al., 2015).

The following HR and HRV measures are used, which are calculated using HeartPy (Gent, 2019):

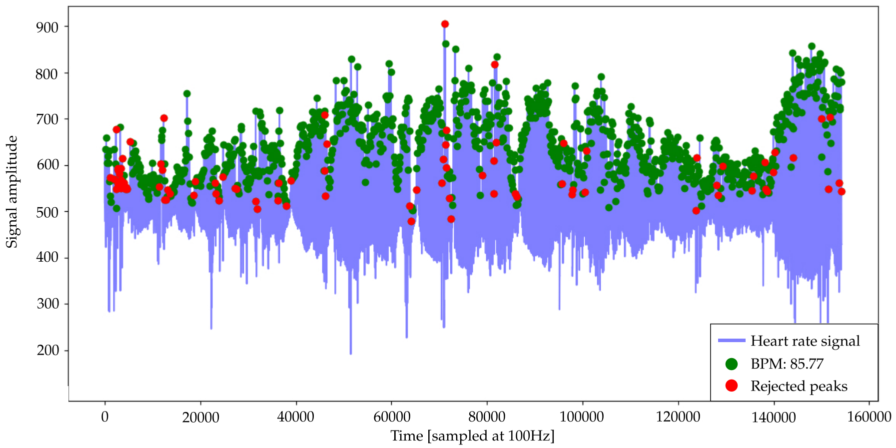
- Average HR [bpm]: calculated from the moment when driving starts to the moment when the vehicle is brought to a full stop at the end of the run.
- The effect of the TOR on HR [bpm]: calculated as the average value of 1 minute following the TOR. This value is compared to a baseline which is the HR before the TOR, which is also calculated as the average value of 1 minute. Moreover, the 1 minute following the TOR is divided into nine 5-second sections to illustrate the heart rate course before and after the TOR (Cho et al., 2015). Increased in HR indicate increases in task demand Mehler et al. (2012).
- Average RMSSD [ms]: calculated from the moment when driving starts to the moment when the vehicle is brought to a full stop at the end of the run.
- The effect of the TOR on RMSSD [ms]: calculated over 1 minute of data following the TOR (Munoz et al., 2015). A baseline value is obtained by applying the same method over 1 minute of data prior to the TOR. The data is not divided into 5-second segments, as RMSSD is not valid for such short durations (the minimum duration for valid measurements is an average over three 10-second durations) (Munoz et al., 2015). Decreases in RMSSD indicate increases in task demand Mehler et al. (2012).
- Average SDNN [ms]: calculated from the moment when driving starts to the moment when the vehicle is brought to a full stop at the end of the run.
- The effect of the TOR on SDNN [ms]: calculated over 1 minute of data following the TOR (Munoz et al., 2015). A baseline value is obtained by applying the same method over 1 minute of data prior to the TOR. The data is not divided into 5-second segments, as SDNN is not valid for such short durations (the minimum duration for valid measurements is an average over three 10-second durations, or a 30-second duration) (Munoz et al., 2015). Decreases in SDNN indicate increases in task demand Mehler et al. (2012).



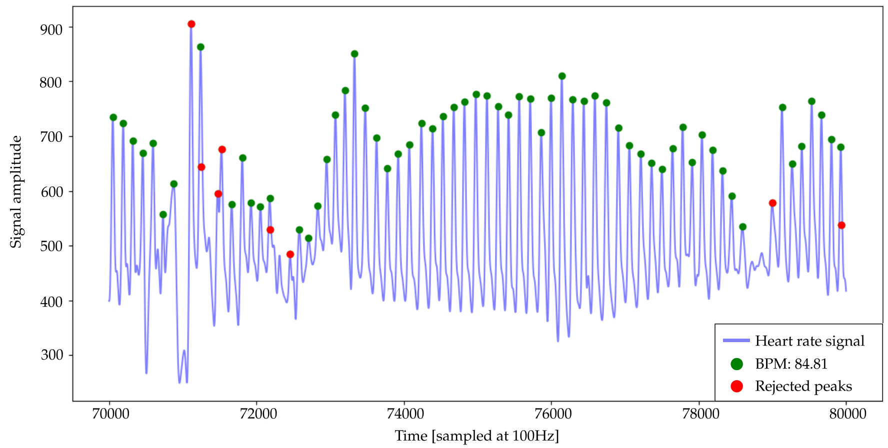
(a) Raw HR data of run 7



(b) Zoomed in raw HR data of run 7



(c) Filtered HR data of run 7



(d) Zoomed in filtered HR data of run 7

Figure 3.8: Example of raw HR data and effect of filtering with a low-pass Butterworth filter

- HF power (ms): calculated from the RRI of the heartbeat in the 0.04 to 0.15Hz spectrum (Task Force of the European Society of Cardiology and Electrophysiology, 1996; Shakouri et al., 2018). A reduction in HF is associated with increasing task demand Mehler et al. (2012). This metric is calculated as the average value of 1 minute of data prior and following the TOR. Again, the data is not divided into smaller sections.

Driving performance

Multiple variables are collected by the driving simulator during the experiment, see Table 3.5. Using these data, the following performance metrics will be calculated:

- TOrt [s]: take-over reaction time is used to determine how long it takes for the driver to shift attention to the driving task. It is measured as the time between the TOR and the moment when the first manoeuvre is executed, this is defined as: the moment when the button in the driving simulator is pushed to shift from automated to manual driving when the available time budget is sufficient, or the first time the steering wheel or pedals are used when the time budget is exceeded (Bueno et al., 2016). Take-over reaction time (TOrt) is an important performance metric to gain insight into the safety aspect of take-overs. If the TOrt exceeds the available time budget, it means that the vehicle is not driven by neither the driver nor the ADS. In reality, however, the ADS may have performed a minimal risk manoeuvre to prevent an accident.
- SD speed [km/h]: provides information about the longitudinal control ability of the driver. Less speed variation indicates a better control ability of the driver.
- SDLP [cm]: provides information about the lateral control ability of the driver. Again, smaller values for SDLP indicate better control ability of the driver.

Matlab scripts used for determining aforementioned driving performance indicators are available via 4TU, DOI: 10.4121/13102763.

Secondary task

As secondary task performance measure, the following measure is used:

- Tetris score: the Tetris score is used to analyse driver distraction from the driving task. Higher scores provide an indication of greater secondary task engagement and therefore a greater distraction from the driving task.

Statistical analysis

The $N = 1$ study design combined with the short experiment duration, possibly time-on-task and learning effects could be found which bias the workload measurements. To determine whether these effects have occurred, trend analysis will be performed on the various workload measurements. Linear workload trends are determined for all runs combined, but will also be determined for the runs of the various days, half-days and parts. Parts are defined as a series of consecutive runs with no breaks in between. Linear trends are obtained by the least squares method and provide insight into the average increase or decrease per run in the workload measurements. A linear regression will be performed over all runs to determine whether the found trend significantly differs from zero. Significance is assumed if $p < 0.05$, the trend is then assumed to be significantly different from zero, thus presumably time-on-task or learning effects have played a role in the workload measurements.

With the aim of this study to provide an empirically validated set of variables for the future study regarding personality, statistical analysis of the workload measures is found of added value. By not only being able to demonstrate differences in workload measurements, but also to be able to demonstrate significant differences, it reinforced the indication of finding workload differences in the future study regarding personality (with over 100 participants). However, it is expected that because of the limited number of observations (due to the

$N = 1$ study design) and the expected large SDs of the workload measurements, significant differences will only be found to a limited extent. Tests for significant differences between attribute levels of the design variables (i.e. between the 0-, 5-, 10- and 15-second time budget) are performed by means of the one-way ANOVA or the Kruskal Wallis test when the residuals are not normally distributed. Checking for normally distributed residuals is done by a face validation of the the QQ-plot of the residuals. Levene's test is performed to check for equal variances, significance is assumed if $p < 0.05$. In case this cannot be assumed, the Welch and Borwn-Forsythe test are used instead of the ANOVA. As the same data set is analysed multiple times a Bonferroni correction is applied which lowers the significance level to $p < 0.0125$. The Bonferroni divides the significance level ($p < 0.05$) by the number of statistical tests performed on the same data (four tests to analyse the effect of time-budget, traffic density, location of the take-over, and NDRT on workload).

3.2.9 Risks and limitations of this study

A limitation of using driving simulator study is that it not fully represents the real world for multiple reasons: graphical representation may distort the drivers perception and behaviour, there is no risk involved for the driver (the participant might show behaviour which deviates from their real behaviour) and there is no or little feedback in the form of sound or sense of movement. Therefore, some argue that results of a driving simulator study can not be generalised to a real traffic situation. However, there are also advantages of a driving simulator study over using real vehicles. For instance, it is possible to control every condition, enabling standardisation of driving scenarios and reproduction of the research data (De Winter et al., 2012). Another advantage is the ease of data collection of a driving simulator. Where in a real-world experiment weather can influence data collection, this cannot hurt the data collection in a driving simulator. Lastly, the participant in a driving simulator will never be physically at risk, whereas this is not the case for experiments in real-vehicles. Despite all the pros and cons, since this study included a $N = 1$ self-experiment during COVID-19, conducting a driving simulation experiment was the only option. Namely, TOR can be simulated in a real world driving environment, but this is done through a Wizard of Oz experiment where the participant is tricked during automated driving as the researcher or assistant is driving the vehicle during automated driving. This was certainly not possible with an N = 1 self-experiment, making it the only logical choice to conduct a driving simulator experiment.

Another risk of this study, is the use of HR and HRV indices as workload measure. HR and HRV are relatively simple to measure, but interpretation and analysis of the results can be difficult. HRV can reflect the participants task-related effort, but can also reflect the participant's physical state, such as blood pressure and respiratory rate. Moreover, both HR and HRV are found to reflect a time-on-task effect (De Winter et al., 2014; Heikoop et al., 2019a). As participants get more accustomed to the experiment and associated decreases in driver vigilance, the HR drops, and the HRV increases. Therefore, interpretation of an increase or decrease in these measures can be difficult, as it can be attributed to multiple factors. However, by combining HR and HRV with the RTLX for subjective workload and driving and secondary task performance measures, increases or decreases in HR and HRV could be attributed to other confounding factors instead of variations in workload. In addition, by conducting the simulation runs in randomised order, the time-on-task effect on workload measurements can be reduced (Section 3.2.3).

Data collection issues could pose a problem, however, as various workload measurements will be included, it is expected that a data collection issue in one of the measurements will not be detrimental for the study. In addition, as the finger sensor is sensitive to record ambient noise to such an extent that the recording becomes unusable, a chest strap will also be used to record physiological workload during the experiment. However, as the chest strap logs data at only 1Hz, the data can only be used as substitute to the mean HR and not for all other physiological workload measures.

A limitation to the statistical analysis methodology is the increased probability of finding Type II errors, where a non-significant difference is not rejected (i.e. increased probability of a false-negative). However, by applying a Bonferroni correction which reduces the statistical significance level from 5% to 1.25%, the probability of Type II errors.

Another limitation of this study is the design of the driving simulation itself. Namely, some design variables varied in the simulation runs are already revealed to the participant at the start of the experiment. e.g. The traffic density at the start of the experiment is the same as the traffic density at the TOR location. Therefore possibly, TOR-induced workload will be less in this experiment compared to an experiment which only reveals all relevant design variables at the TOR location. Moreover, the used take-over modality will probably need some reconsideration in future experiments, as it is preferred to allow overruling of the ADS by pressing the driving pedals or pressing a button on the steering wheel.

4 RESULTS OF THE EXPERIMENT

In this chapter, the results of the driving simulation study are presented. First, the results of the selection procedure for participating in the experiment regarding the role of personality are presented in Section 4.1. Descriptive statistics of the applicants and the 100 participants if the experiment was continued are provided. After that, the results of the validation study are presented. In Section 4.2 the measurements of the different workload measures are validated. After that, the effect on workload of the four task load conditions that were varied in the driving simulation scenarios are presented. First, the effect of the four time budgets is presented in Section 4.3. Second, in Section 4.4 the effect of the different traffic densities on workload is presented. Third, the effect on workload of the location of the TOR is presented in Section 4.5. Fourth and last, the effects of the two NDRTs on workload are presented in Section 4.6.

4.1 PERSONALITY EXPERIMENT: PARTICIPANT SELECTION

A total of 159 people applied for participating in the driving simulation experiment regarding personality and automated driving. 58 women applied and 101 men with an average age of 46.2 years. 138 applicants have the Dutch nationality and 21 a foreign nationality. On average, they have held a driver's license for 26.5 years. 35.0% of the applicants drive daily, 20.1% four to six days a week, 22.6% one to three days a week, 10.0% once a week, 10.7% less than once week, and 1.6% never drive. 52.2% of the applicants had no previous experience with any Automated Driving System (ADS). The dominant personalities of the applicants are illustrated in Figure 4.1. Neuroticism is the most represented dominant personality trait, with 28.9% of the applicants having this trait as the most dominant trait. Openness is the least represented dominant personality trait, with 14.5% of the applicants having openness as dominant personality trait. Extraversion, agreeableness and conscientiousness are almost equally represented, with a distribution of, respectively, 17.6%, 19.5%, and 19.5%.

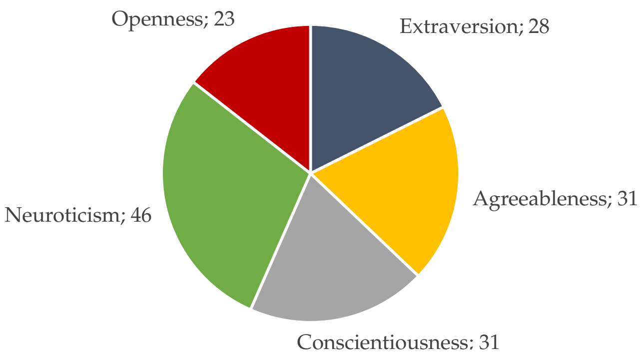


Figure 4.1: Distribution of dominant personality of the applicants, N = 159

For each personality trait, the 20 highest scoring participants are selected for the experiment. This resulted in the selection of 100 participants, consisting of 57 men and 43 women, with an average age of 45.9 years. 86 participants held the Dutch nationality and 14 a foreign nationality. On average, they have held a driver's license for 26.3 years. Of the selected participants, 35.0% drive daily, 20.0% four to six days a week, 28.0% one to three days a week, 7.0% once a week, 9.0% less than once week, and 1.0% never drive. 54.0% of the participants had no previous experience with any ADS. Table 4.1 provides insight into the dominant personality trait scores of the participants selected for the experiment. No participant achieved the maximum possible score for the neuroticism trait, with a maximum score of 36 points instead of the possible 40 points.

Table 4.1: Distribution dominant personality traits of selected participants of the experiment

Dominant personality trait	N	Max. possible score	Max. obtained score	Mean	SD
Extraversion	20	40	40	36.55	2.04
Agreeabless	20	45	45	41.55	2.16
Conscientousness	20	45	45	41.85	2.08
Neuroticism	20	40	36	28.15	3.05
Openness	20	50	50	43.80	3.47

4.2 PATTERNS IN WORKLOAD MEASUREMENTS

The experiment has been performed on five alternating days for which a 5-day schedule was proposed (Table 3.4). Ultimately, this schedule was not adhered to as initial tests on the first day of the experiment revealed issues with the road network in the scenarios that led to vehicle crashes and stuttering of the scenario. Eventually, on the third day available for the experiment, the driving simulation experiment was started. In total, 64 scenarios were prepared for the experiment with different task load conditions. However, 72 runs have been performed for 63 scenarios, as one scenario was accidentally tested twice and eight scenarios have been rerun due to expected data collection issues after a face validation of the data. After the validation of the collected data after completion of the experiment, it was found that nine scenarios still had incorrect or missing finger sensor data, therefore for these scenarios missing data is reported in the analysis. No missing or incorrect data is reported for the Raw Task Load Index (RTLX), chest strap, driving metrics and Tetris. Table 4.2 provides an overview of the scenarios that encountered data collection issues.

Table 4.2: Scenarios with data collection issues, if a scenario is rerun, the run number is listed

Scenario	Issue	Rerun #	Scenario	Issue	Rerun #
4	Finger sensor	-	23	Face validation	66
5	Finger sensor	-	30	Finger sensor	-
6	Finger sensor	-	31	Finger sensor	-
10	Finger sensor	72	32	Finger sensor	-
11	Finger sensor	67	35	Finger sensor	-
12	Finger sensor	65	36	Finger sensor	-
13	Finger sensor	69	37	Finger sensor	-
14	Finger sensor	70	44	Face validation	71
16	Not ran*	-	45	Face validation	68

*Instead, scenario 54 was run, for its characteristics see Table 3.3

On average, a run took 3:46 minutes to complete, adding up to a total of 4:31 hours of driving time to complete all 72 runs. On Day 1 of experimenting 37 runs have been completed, 27 runs on Day 2 and eight runs on Day 3. The final eight runs were reruns as incorrect finger sensor data was expected for eight scenarios. During the 72 runs, no accidents have occurred that would have led to early termination of the run.

In this section, the data which is collected during the experiment is validated, and workload trends are obtained. When a workload measure is found invalid, the measure is not used in the assessment of the effect of the different design variables (time budget, traffic density, location of the take-over request (TOR), and the Non-Driving Related Task (NDRT)) on workload). The workload trends were obtained by analysing the 72 runs in ascending order. This way, workload was analysed for the different days and half-days (morning and afternoon). Experiment parts are distinguished, which are defined as a series of runs that are performed without breaks in between. The linear workload trend provides insight into the average trend by which the workload score on average increases or decreases per run. The method of determining workload trends is explained in Figure 3.2.8. The linear workload trend provides insight into the average increase or decrease of the workload measurement per run as the experiment progressed. Figures are provided illustrating the workload scores per run in ascending order, besides the linear workload trend also polynomial trends are illustrated. The polynomial trends are included for illustrative purposes as a distinction of the different parts. Tests for significance are done by performing a simple linear regression analysis over all 72 runs, in order to test whether the found trends are significantly different from zero.

4.2.1 Subjective workload

The results of the RTLX are used for assessing the subjective workload that was experienced during the simulation. For this, the workload scores on the six RTLX scales (mental demand, physical demand, temporal demand, performance, effort, and frustration) are presented, as well as the overall workload score which is obtained by averaging the scores of the six workload scales. Figure 4.2a illustrates the overall workload for runs 1 to 72. The workload as measured on the six RTLX scales is illustrated in Figure 4.3a to 4.3f. The scale ranges from low (0%) to high (100%), except the performance scale that ranges from good (0%) to poor (100%). See Table 3.2.8 for definitions of the scales.

Overall workload

All runs combined, an overall workload score of 31.12% and a trend of +0.04% is measured (Table 4.3). By means of a simple linear regression this trend is analysed and was found to be non-significant ($F(1, 70) = 0.168$ and $p = 0.683$). The runs on Day 1 show a negative trend of -0.71%, which is due to a large difference in mean overall workload between the morning and afternoon runs. For Day 2, a positive trend of +0.28% is found. From Figure 4.2a it can be seen that in part 8 (the first part of Day 2) lower overall workload is reported compared to the other parts of that day. The runs of part 9 report a relatively higher overall workload score. With regards to Day 3, a positive linear trend of +0.22% is found.

Table 4.3: Trend analysis RTLX overall workload, scores range between low (0%) and high (100%)

	Overall workload [%]		
	N	**Mean(*SD*)**	**Trend**
All runs	72	31.12(*16.98*)	+ 0.04
Day 1	37	27.97(*17.34*)	- 0.71
Morning	16	37.97(*16.50*)	+ 0.11
Afternoon	21	20.36(*14.01*)	- 0.11
Day 2	27	35.34(*17.39*)	+ 0.28
Morning	19	36.40(*20.25*)	+ 1.19
Afternoon	8	32.81(*7.62*)	+ 1.22
Day 3	8	31.46(*11.68*)	+ 0.22
Morning	8	31.46(*11.68*)	+ 0.22

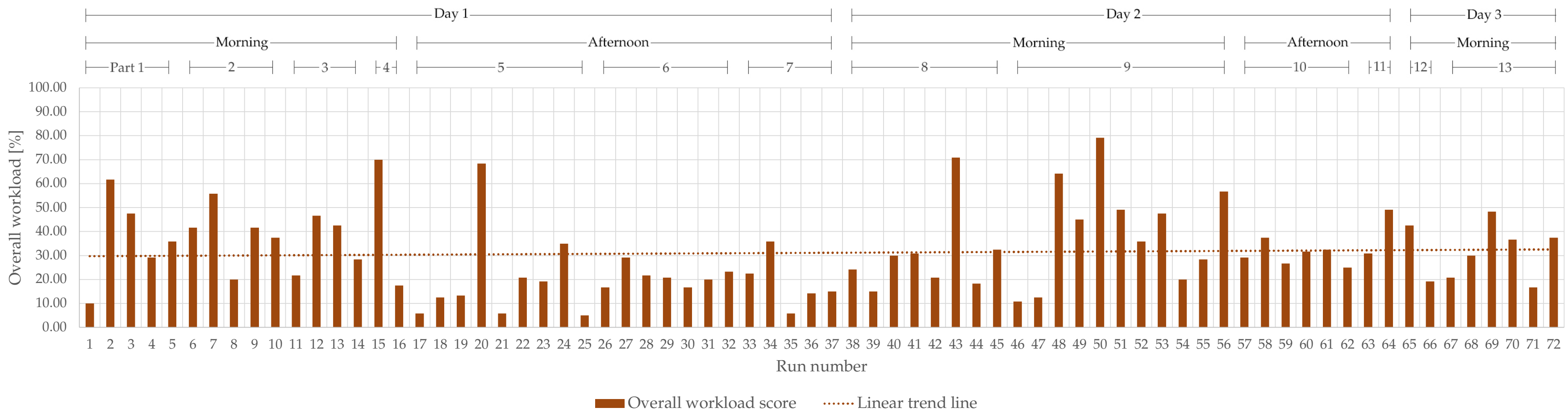
Workload on the six RTLX scales

Despite the scale ranging from 0% to 100%, almost all workload scores on the different scales of the RTLX report values below 50% (Table 4.4). The effort and temporal demand scales measure a relatively high workload compared to the other scales, with an average workload of 37.78% and 37.22%, respectively. The performance scale reports the lowest average workload over all 72 runs of 19.44%. No scale reports a constant positive or negative trend over all days. Therefore the linear trend over all runs is small, frustration reported the largest trend of +0.12% over all runs. A linear regression was performed, and found no significant trends: $F(1,70) = 0.058$ with $p = 0.811$ for mental demand, $F(1,70) = 0.103$ and $p = 0.750$ for physical demand, $F(1,70) = 0.093$ and $p = 0.761$ for temporal demand, $F(1,70) = 0.128$ and $p = 0.722$ for performance, $F(1,70) = 0.173$ and $p = 0.679$ for effort, and $F(1,70) = 1.007$ and $p = 0.319$ for frustration.

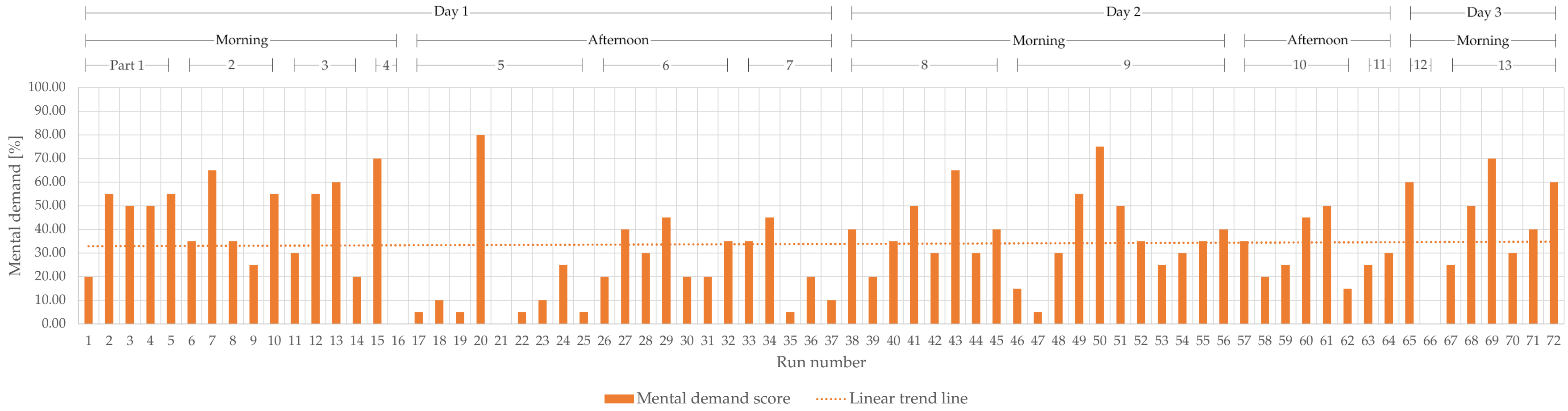
Table 4.4: Trend analysis RTLX scales. Scores range between low (0%) and high (100%), except the performance scale that ranges from good (0%) to poor (100%).

		Mental demand [%]		Physical demand [%]		Temporal demand[%]	
	N	**Mean(*SD*)**	**Trend**	**Mean(*SD*)**	**Trend**	**Mean(*SD*)**	**Trend**
All runs	72	33.82(*19.78*)	+ 0.03	29.51(*10.05*)	- 0.03	37.22(*28.72*)	+ 0.05
Day 1	37	31.08(*21.67*)	- 0.76	26.89(*21.32*)	- 1.08	33.24(*26.62*)	- 0.69
Morning	16	42.50(*19.49*)	- 0.60	41.56(*20.87*)	- 0.39	40.63(*23.93*)	+ 1.51
Afternoon	21	22.38(*19.40*)	+ 0.28	15.71(*13.72*)	- 0.10	27.62(*27.72*)	- 1.58
Day 2	27	35.19(*15.60*)	- 0.26	33.15(*16.12*)	+ 0.21	44.26(*32.51*)	+ 0.24
Morning	19	37.11(*16.78*)	+ 0.14	33.95(*18.97*)	+ 0.85	45.00(*33.21*)	+ 0.67
Afternoon	8	30.63(*12.08*)	- 0.42	31.25(*5.82*)	+ 1.67	42.50(*32.95*)	+ 4.88
Day 3	8	41.88(*22.98*)	+ 2.80	29.38(*17.00*)	+ 0.89	31.88(*22.02*)	+ 2.08
Morning	8	41.88(*22.98*)	+ 2.80	29.38(*17.00*)	+ 0.89	31.88(*22.02*)	+ 2.08

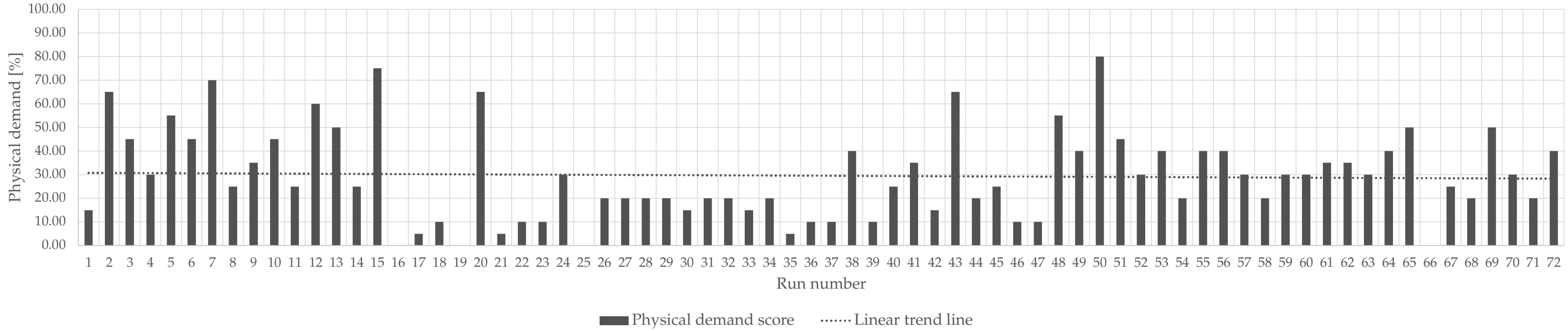
		Performance [%]		Effort [%]		Frustration [%]	
	N	**Mean(*SD*)**	**Trend**	**Mean(*SD*)**	**Trend**	**Mean(*SD*)**	**Trend**
All runs	72	19.44(*14.81*)	+ 0.03	37.78(*21.06*)	+ 0.05	28.96(*20.38*)	+ 0.12
Day 1	37	16.35(*14.08*)	- 0.61	35.68(*22.27*)	- 0.38	24.59(*18.27*)	- 0.75
Morning	16	24.38(*16.01*)	- 0.21	43.75(*22.99*)	+ 0.49	35.00(*18.62*)	- 0.16
Afternoon	21	10.24(*8.58*)	- 0.17	29.52(*20.12*)	+ 0.79	16.67(*13.72*)	- 0.05
Day 2	27	25.19(*15.84*)	+ 0.37	40.93(*21.26*)	+ 0.59	33.33(*22.62*)	+ 0.52
Morning	19	26.05(*17.53*)	+ 1.50	40.79(*24.28*)	+ 1.57	35.53(*25.81*)	+ 2.41
Afternoon	8	23.13(*11.63*)	- 0.54	41.25(*12.75*)	+ 0.92	28.13(*11.93*)	+ 0.77
Day 3	8	14.38(*8.21*)	- 1.01	36.87(*14.37*)	- 3.75	34.38(*19.72*)	+ 0.30
Morning	8	14.38(*8.21*)	- 1.01	36.87(*14.37*)	- 3.75	34.38(*19.72*)	+ 0.30



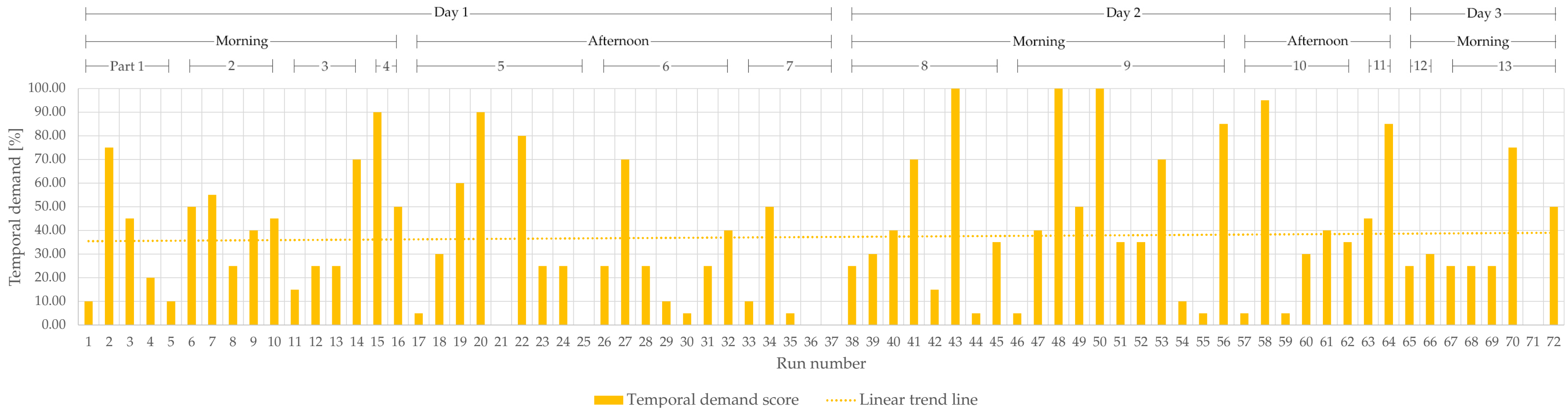
(a) Overall workload



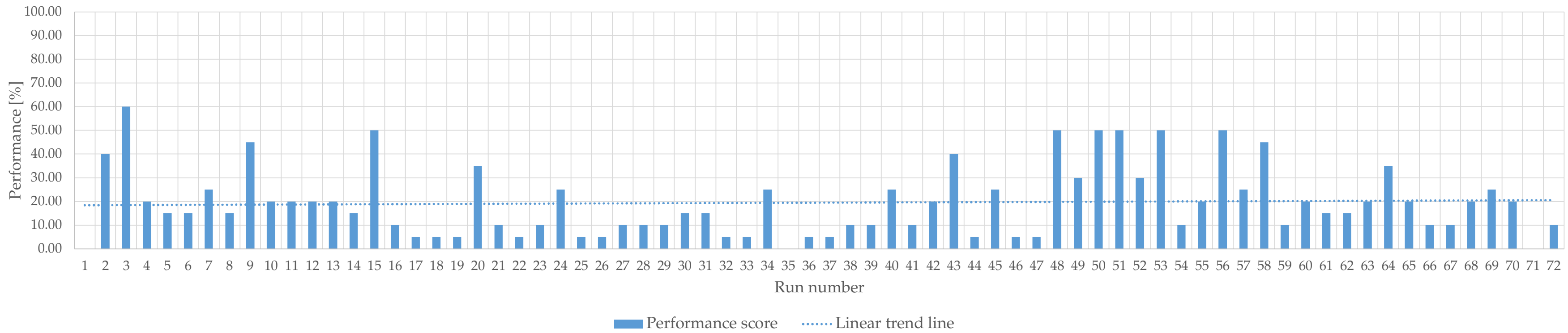
(a) Mental demand



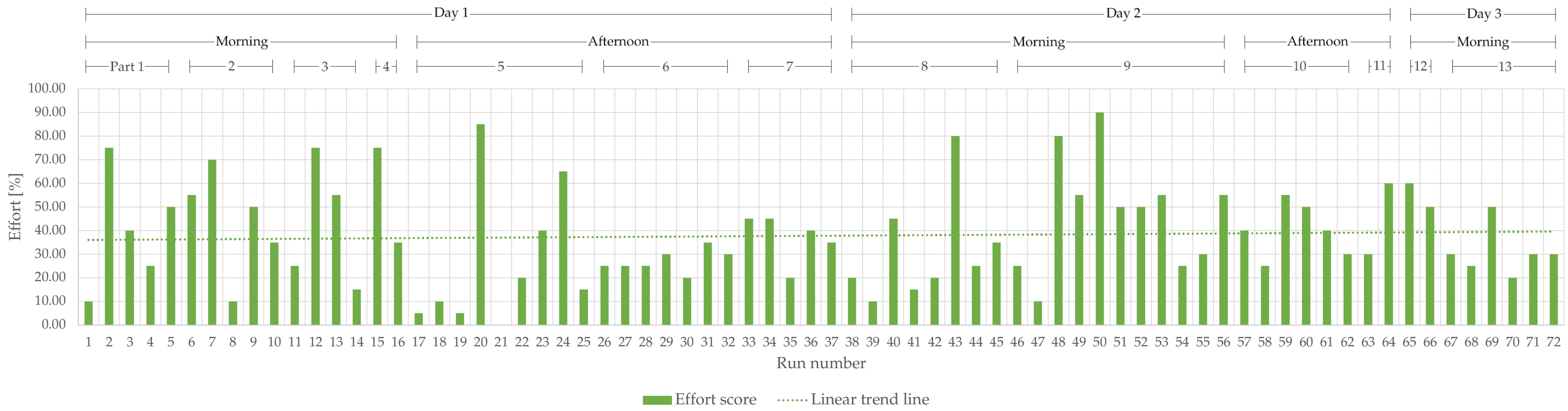
(b) Physical demand



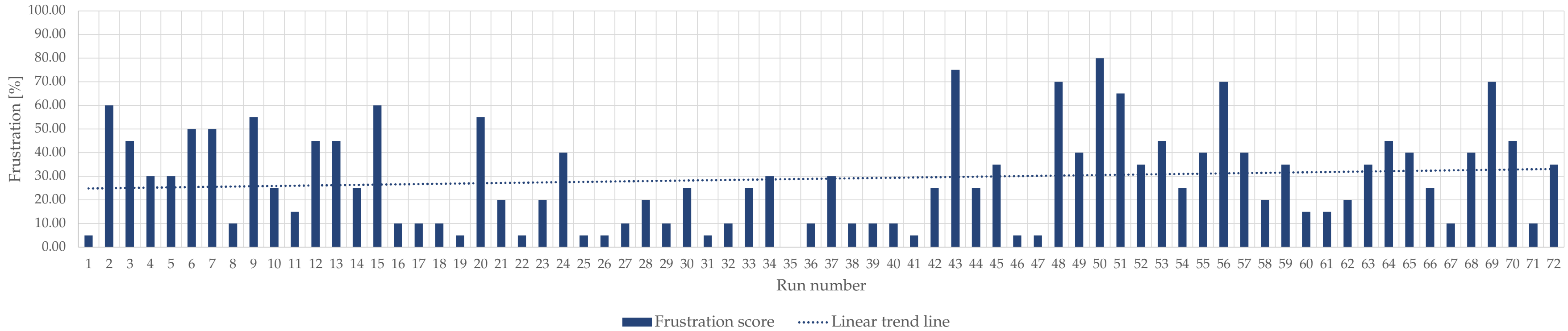
(c) Temporal demand



(d) Performance



(e) Effort



(f) Frustration

Figure 4.3: Overall workload and scores on the six RTLX scales for all 72 runs performed during the experiment

4.2.2 Physiological workload

The results of the physiological workload measurements are presented in this subsection. Both the mean Heart Rate (HR), Root Mean Square of Successive Differences (RMSSD), and Standard Deviation of Normal to Normal intervals (SDNN) during the entire experiment are presented as well as the effect of the TOR on these measures. These measures are elaborated in Section 3.2.8. Figure 4.4 to 4.6 illustrate the mean physiological workload during the experiment and Figure 4.7 illustrates the effect of the TOR on these measures. The found results are tabulated for the different days and half-days in Table 4.6.

During the experiment, the driver wore a finger sensor and chest strap to measure physiological workload. Chest strap data was used when the finger sensor data deviated more than 3 beats per minute (bpm) (after the data filtering) from the mean HR measured by the chest strap. Due to the low sampling rate of the chest strap (Section 3.2.5), the chest strap data is only used as substitute for the mean HR. Missing data is therefore reported for the mean RMSSD, SDNN and effect of the TOR on HR, RMSSD and SDNN.

Heart Rate

Figure 4.4 illustrates the HR as it fluctuates between run 1 and run 72. Chest strap data is used for some runs that encountered finger sensor data collection issues (Table 4.2). The finger sensor data is illustrated by a circle and chest strap data by a cross. In total, 14 runs encountered issues with the collected finger sensor data. The mean HR and linear workload trends by which the mean HR fluctuates are presented in Table 4.7a. The average highest HR was measured in the afternoon on Day 1, which is caused by the elevated HR in Part 5 (the first part of the afternoon). During the experiment, mean HR decreased on average by - 0.10 bpm per run. This trend was found significant by a simple linear regression analysis ($F(1, 70) = 6.042$ and $p = 0.016$).

The increases and decreases in HR after the TOR are illustrated in Figure 4.7a and tabulated in Table 4.7b. Runs 21 and 23 report a close to zero effect on HR of, respectively, +0.04 and +0.01 bpm. The greatest decreases in HR are reported in Runs 52 and 66, respectively with a decrease of -6.57 and -7.70 bpm after the TOR. The largest increases in HR are reported in runs 8 and 38, with an increase of +6.10 and +6.59 bpm, respectively. The large average effect of the take-over in the morning runs of Day 1 of +2.83 bpm stands out, as on average an effect of +0.85 bpm is reported. A linear regression analysis was performed and found a non-significant trend in the effect of the TOR on HR ($F(1, 56) = 1.321$ and $p = 0.255$).

RMSSD (Heart Rate Variability)

Table 4.7a and Figure 4.5 present the average RMSSD measured in the experiment. Missing data is only reported for runs on Day 1, 14 runs in total, of which 8 in the morning and 6 in the afternoon. The RMSSD shows a similar pattern as the mean HR; a relative large difference is reported for runs in the morning and afternoon of Day 1, with a difference of 15.91 ms between the half-days. A mean RMSSD of 37.17 ms was measured over all runs (Table 4.7a). This meets the norm with a mean of 42 ms and Standard Deviation (SD) of 15 ms for short-term measurements (Shaffer and Ginsberg, 2017). A trend of +0.15 ms is found over all runs. A linear regression analysis was performed, but found a non-significant effect ($F(1, 56) = 1.765$ and $p = 0.189$).

Figure 4.7b illustrates the effect of a the TOR on RMSSD. The effect on RMSSD varies widely during the runs, from a decrease of -20.93 ms in Run 34 to an increase of +43.44 ms in Run 63. On average, the TOR resulted in an increase of + 0.53 ms per run. A positive trend of + 0.10 ms per run was measured, but found non-significant by performing the linear regression analysis ($F(1, 56) = 2.069$ and $p = 0.156$).

SDNN (Heart Rate Variability)

The average SDNN during the 72 runs is presented in Table 4.7a.The SDNN measured on average 37.19 ms, which meets the norm with a mean of 50 ms and a SD of 16 ms for short-term measurements (Shaffer and Ginsberg, 2017). The average SDNN of run 1 to 72 show a similar pattern as the RMSSD values, as can be seen in Figure 4.5 and Figure 4.6. Over all runs a trend of +0.12 ms is found. This trend is analysed by means of a linear regression analysis and was found insignificant($F(1, 56) = 2.889$ and $p = 0.095$).

The linear trend of all runs and daily and half-day linear trends by which the SDNN increases or decreases are presented in Table 4.7b. Figure 4.7b illustrates the effect of a TOR on the SDNN. It shows that the SDNN varies widely during the runs, from a decrease of -59.03 ms in Run 15 to an increase of +27.37 ms in Run 63. On average, the SDNN decreased by -1.91 ms after the TOR. The effect on SDNN increased on average by +0.17 ms per run. This trend was found significant ($F(1, 56) = 4.040$ and $p = 0.049$) by means of a linear regression analysis.

HF power (Heart Rate Variability)

An average High-Frequency (HF) power of 21,984.22 ms^2 was found over all runs. This value lies well outside the norm for short-term measurements with a mean of 657 ms^2 and SD of 777 ms^2 (Shaffer and Ginsberg, 2017). Only in one run a HF is found within the norm: run 18 with a HF power of 1,058.14 ms^2. As for too little runs a reliable HF power is found, HF power will not be used as workload measure.

Table 4.6: Trend analysis physiological workload

(a) Mean physiological workload

	HR			RMSSD [ms]			SDNN [ms]	
	N(*)	**Mean(*SD*)**	**Trend**	**N(*)**	**Mean(*SD*)**	**Trend**	**Mean(*SD*)**	**Trend**
All runs	72(0)	87.33(*7.48*)	- 0.10*	72(14)	37.62(*14.04*)	+ 0.15	37.19(*11.32*)	+ 0.12
Day 1	37(0)	89.39(*8.73*)	+ 0.06	37(14)	35.33(*18.32*)	- 0.01	34.17(*13.94*)	- 0.10
Morning	16(0)	83.92(*6.60*)	- 1.32	16(8)	47.24(*21.76*)	+ 3.64	43.15(*18.16*)	+ 2.38
Afternoon	21(0)	93.55(*7.91*)	- 1.17	21(6)	28.97(*12.89*)	+ 2.18	29.75 (*8.35*)	+ 1.26
Day 2	27(0)	84.89(*4.44*)	- 0.13	27(0)	37.57 (*9.38*)	+ 0.07	38.32 (*8.44*)	+ 0.04
Morning	19(0)	85.56(*4.11*)	- 0.39	19(0)	38.70 (*7.41*)	+ 0.40	39.38 (*8.19*)	+ 0.34
Afternoon	8(0)	86.68(*5.37*)	- 1.97	8(0)	34.87(*13.19*)	+ 4.28	35.80 (*9.03*)	+ 3.23
Day 3	8(0)	82.70(*6.92*)	- 2.69	8(0)	44.45(*12.37*)	+ 4.44	42.03(*10.35*)	+ 3.75
Morning	8(0)	82.70(*6.92*)	- 2.69	8(0)	44.45(*12.37*)	+ 4.44	42.03(*10.35*)	+ 3.75

*Significant at the 5% level

(b) TOR-induced physiological workload, computed as the difference between 1 minute measurements before and after the TOR.

		HR		RMSSD [ms]		SDNN [ms]	
	N(*)	**Mean(*SD*)**	**Trend**	**Mean(*SD*)**	**Trend**	**Mean(*SD*)**	**Trend**
All runs	72(14)	+ 0.85(*2.88*)	- 0.02	+ 0.45(*11.60*)	+ 0.10	- 1.91(*13.13*)	+ 0.17*
Day 1	37(14)	+ 1.01(*2.41*)	- 0.06	- 2.38(*9.97*)	+ 0.05	- 6.37(*15.16*)	- 0.07
Morning	16(8)	+ 2.83(*2.18*)	+ 0.20	- 3.21(*9.51*)	- 0.64	- 8.70(*23.16*)	- 2.53
Afternoon	21(6)	+ 0.04(*1.96*)	+ 0.20	- 1.93(*10.50*)	- 0.48	- 4.60 (*9.39*)	+ 0.31
Day 2	27(0)	+ 1.04(*3.00*)	- 0.09	+ 1.66(*12.37*)	- 0.11	+ 0.46(*11.24*)	+ 0.04
Morning	19(0)	+ 1.09(*3.16*)	- 0.22	+ 2.04(*9.42*)	+ 0.38	+ 0.89(*10.54*)	+ 0.19
Afternoon	8(0)	+ 0.90(*2.78*)	- 0.20	+ 0.75(*18.42*)	+ 1.28	- 0.56(*13.50*)	+ 1.45
Day 3	8(0)	- 0.25(*3.77*)	+ 0.82	+ 4.49(*12.83*)	- 0.22	+ 2.90 (*9.97*)	+ 1.64
Morning	8(0)	- 0.25(*3.77*)	- 0.81	+ 4.49(*12.83*)	- 0.22	+ 2.90 (*9.97*)	+ 1.64

*Significant at the 5% level

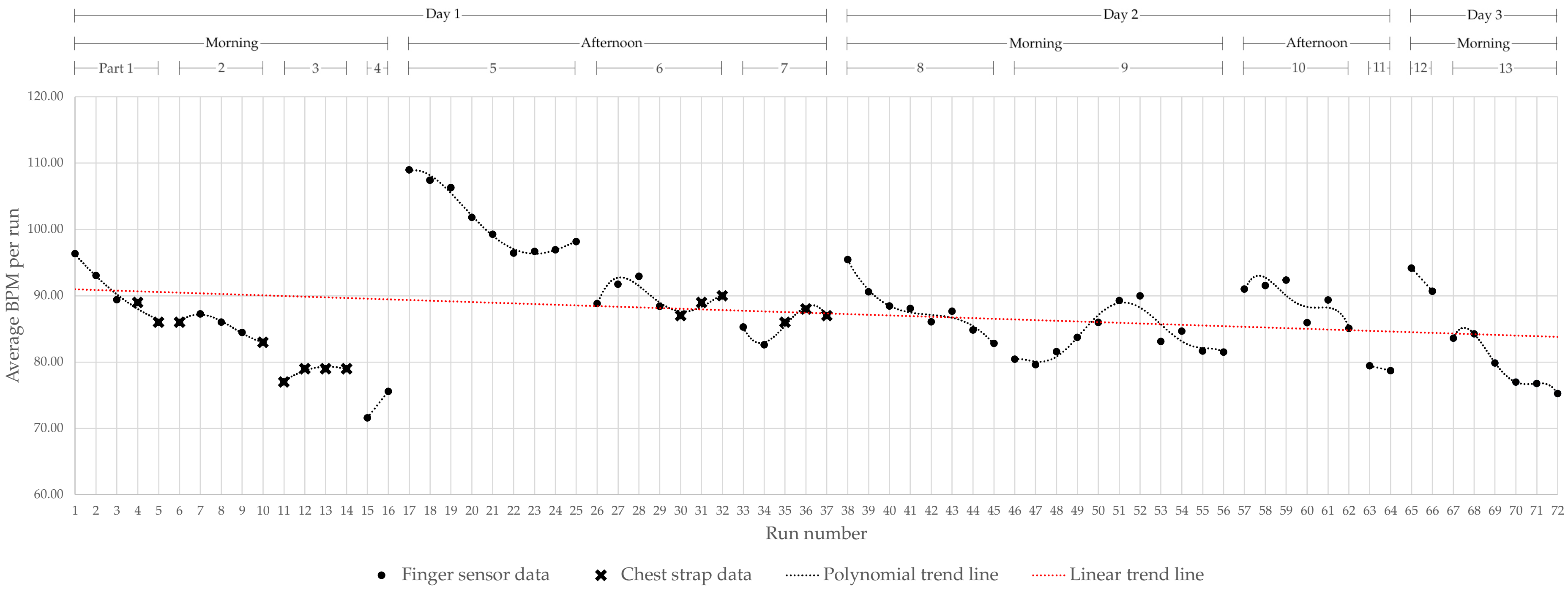


Figure 4.4: Average BPM of the runs performed on the 1st, 2nd and 3rd day of experimenting, measured over the entire run. In case the finger sensor data was faulty, the average HR as measured by the chest strap is used. The polynomial trend lines are included for illustrative purposes for distinction of the different parts.

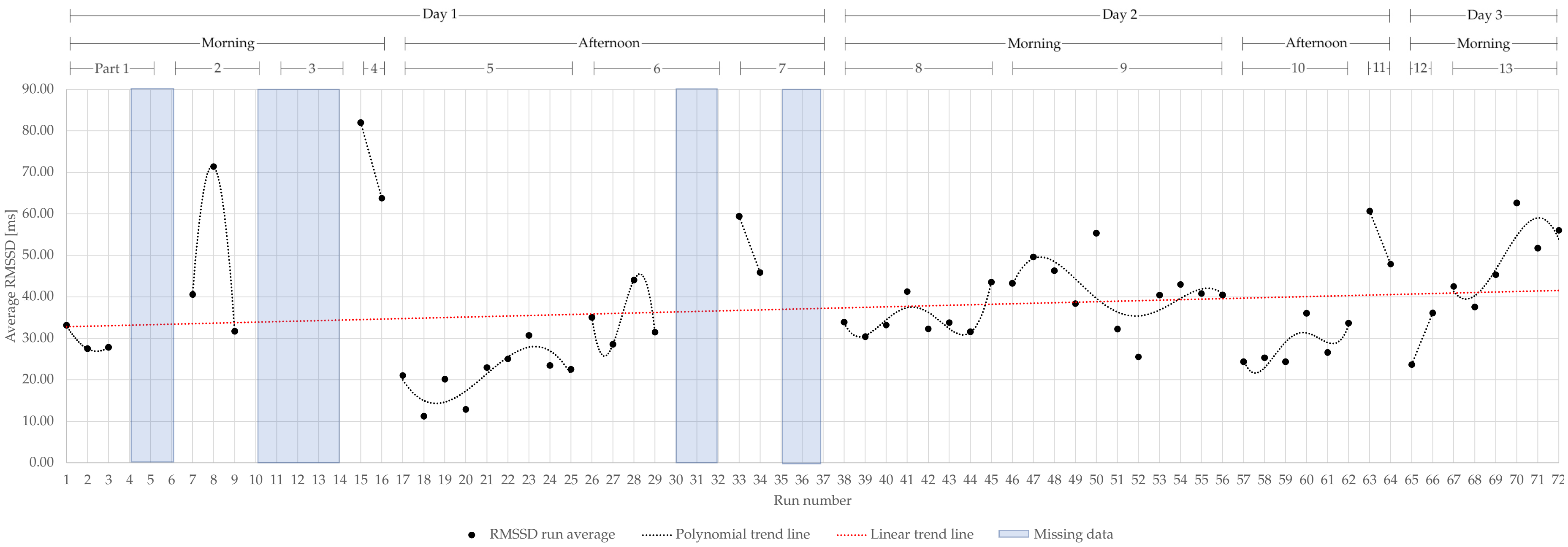


Figure 4.5: Average RMSSD of the runs performed on the 1st, 2nd and 3rd day of experimenting. The polynomial trend lines are included for illustrative purposes for distinction of the different parts.

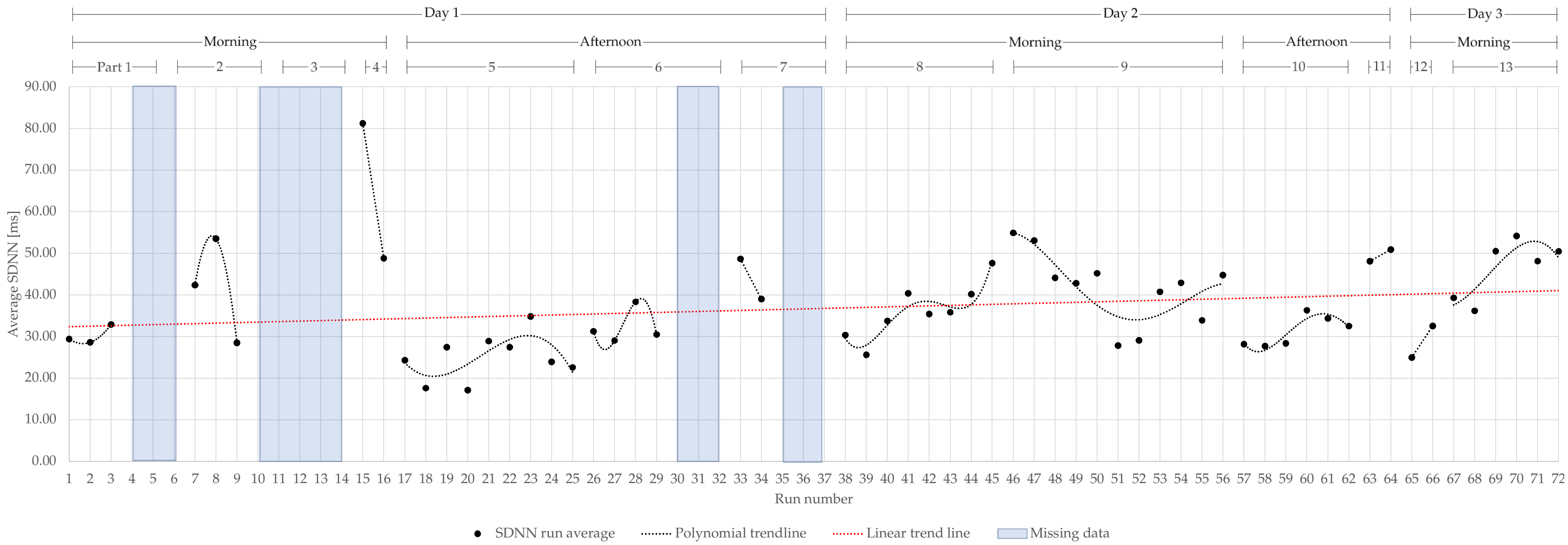
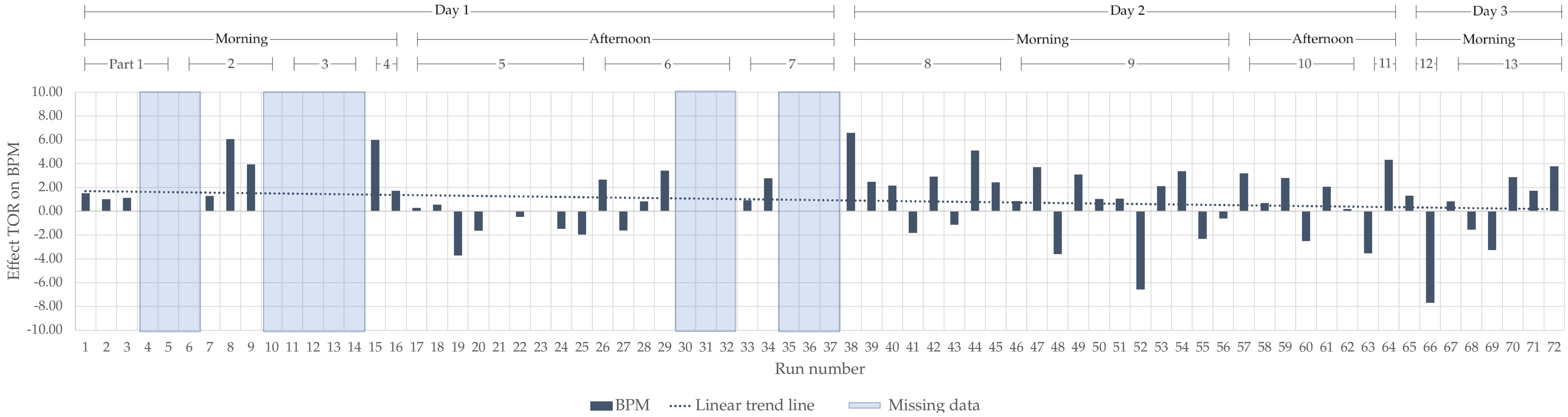
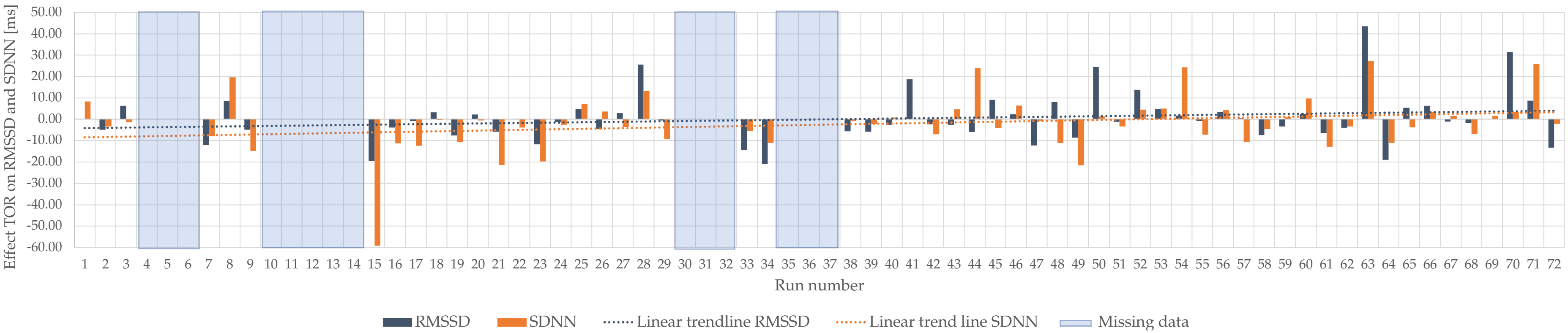


Figure 4.6: Average SDNN of the runs performed on the 1^{st}, 2^{nd} and 3^{rd} day of experimenting. The polynomial trend lines are included for illustrative purposes for distinction of the different parts.



(a) Effect of the TOR on HR.



(b) Effect of the TOR on RMSSD and SDNN

Figure 4.7: Effect of the TOR on physiological workload, computed as the difference between 1 minute before and after the TOR.

4.2.3 Driving measures

Data logged by the driving simulator is used for the driving performance indicators. For this, the average take-over reaction time (TOrt), longitudinal control ability as indicated by the SD speed after the take-over, and lateral control ability as indicated by the Standard Deviation of Lateral Position (SDLP) is analysed. Figure 4.8 to 4.10 illustrate the found driving performance measures for the 72 runs in ascending order. The mean, SDs and found trends are tabulated and can be found in Table 4.9.

Take-over reaction time

Figure 4.8 illustrates the TOrt of the 72 runs. The figure illustrates that as the experiment progressed, the TOrt showed smaller spreading of the TOrts. The average difference between the longest and shortest TOrt on Day 1 measured 9.22 seconds on Day 1, 7.72 seconds on Day 2 and 5.93 seconds on Day 3. See Table 4.8 for an overview of the runs that exceeded the available time budget. There is no clear relation between run number, time of day and the duration of the exceeded time budget. The linear trend of all runs and daily and half-day linear trends by which the TOrt increases or decreases are presented in Table 4.9. On average, the runs on Day 1 report a longer TOrt with an average of 5.74 seconds. The morning runs of Day 1 report the highest average TOrt of all half-days, with an averageTOrt of 6.40 seconds. The shortest TOrts are reported in the morning of Day 2, with an average TOrt of 4.89 seconds. An average TOrt of 5.38 seconds is reported as well as a negative trend of -0.02 seconds. This trend is found insignificant by means of a linear regression analysis ($F(1,70) = 3.647$ and $p = 0.060$).

Table 4.8: Overview of the runs in which the available time budget was exceeded, runs with a 0-second time budget are excluded from this overview.

Run number	Day	Part of day	Exceeded time budget [s]
3	1	Morning	0.08
6	1	Morning	2.15
8	1	Morning	0.20
18	1	Afternoon	1.93
28	1	Afternoon	0.97
34	1	Afternoon	2.65
39	2	Morning	0.08
54	2	Morning	1.35
61	2	Afternoon	1.22
62	2	Afternoon	1.72
72	3	Morning	0.20

Longitudinal control ability

Figure 4.9 illustrates the speed deviation during manual control after the take-over of the 72 runs. The linear trends of all runs and respective days and half-days are presented in Table 4.9. The speed deviations varies, with the highest speed deviation of 10.06 km/h in Run 40 to the lowest speed deviation of 0.99 km/h in Run 9. The speed deviation on Day 3 was smallest with an average of 3.44 km/h, compared to the average speed deviation of 4.96 km/h. The SD speed increased on average by +0.01 km/h. This trend was found insignificant by the linear regression analysis ($F(1,70) = 1.045$ and $p = 0.310$).

Lateral control ability

Figure 4.10 illustrates the SDLP during manual control after the take-over. The linear trends are presented in Table 4.9. Part 8 and 11 stand out by the large spread in SDLP, as such that the smallest and the largest SDLP are both measured in part 8: 7.65 cm in Run 42 and 67.50 cm in run 43. The SDLP increased on average by + 0.01 per run. However, this trend was found insignificant by performing a linear regression analysis ($F(1,70) = 0.028$ and $p = 0.868$).

Table 4.9: Driving performance means and trends over all runs, measured over 1 minute over manual driving after the take-over.

		TOrt [s]		SD speed [km/h]		SDLP [cm]	
	N(*)	**Mean(*SD*)**	**Trend**	**Mean(*SD*)**	**Trend**	**Mean(*SD*)**	**Trend**
All runs	72	5.38(*2.06*)	- 0.02	4.96(*1.98*)	+ 0.01	25.39(*10.33*)	+ 0.01
Day 1	37	5.74(*2.19*)	- 0.04	4.47(*2.01*)	+ 0.02	24.30(*7.66*)	- 0.11
Morning	16	6.40(*2.46*)	- 0.18	4.70(*2.13*)	+ 0.10	26.94(*6.34*)	+ 0.06
Afternoon	21	5.24(*1.86*)	+ 0.10	4.29(*1.95*)	+ 0.16	22.29(*8.11*)	+ 0.37
Day 2	27	4.98(*1.98*)	- 0.01	6.09(*1.37*)	- 0.06	26.93(*14.21*)	- 0.32
Morning	19	4.89(*2.10*)	- 0.04	6.29(*1.42*)	- 0.08	27.47(*15.12*)	- 0.84
Afternoon	8	5.20(*1.77*)	- 0.25	5.84(*1.31*)	- 0.10	25.62(*12.63*)	+ 2.41
Day 3	8	5.04(*1.58*)	- 0.26	3.44(*1.71*)	- 0.23	25.25(*3.73*)	+ 0.97
Morning	8	5.04(*1.58*)	- 0.26	3.44(*1.71*)	- 0.23	25.25(*3.73*)	+ 0.97

4.2.4 Secondary task

Two tasks during automated driving were varied in the experiment, which are the monitoring task and playing Tetris. The scores obtained in the Tetris game were used as secondary task measure. Figure 4.11 illustrates the Tetris scores that were obtained in the experiment runs. The linear trends are presented in Table 4.10. Note that in 37 runs no score is reported as these runs has a monitoring task instead of playing Tetris. The figure and table show that the mean score increases as the experiment progresses, with an average score of 933.20 on Day 1 to an average score of 2143.67 on Day 3, with a linear trend of + 9.92 points. This is also confirmed by a linear regression analysis, which reports a significant difference between the Tetris scores as the experiment progressed, with $F(1,33) = 6.431$ and $p = 0.016$.

Table 4.10: Tetris score trends and averages of all runs.

	N	**Mean(*SD*)**	**Trend**
All runs	35	1105.29(*571.64*)	+ 9.92*
Day 1	11	933.20(*462.88*)	+ 4.72
Morning	8	924.75(*394.59*)	+ 58.05
Afternoon	3	942.86(*563.98*)	- 15.16
Day 2	17	1073.88(*479.87*)	+ 7.37
Morning	10	1089.10(*484.18*)	+ 9.16
Afternoon	7	1052.14(*511.36*)	+ 135.75
Day 3	3	2143.67(*604.51*)	- 238.76
Morning	3	2143.67(*604.51*)	- 238.76

*Significant at the 5% level

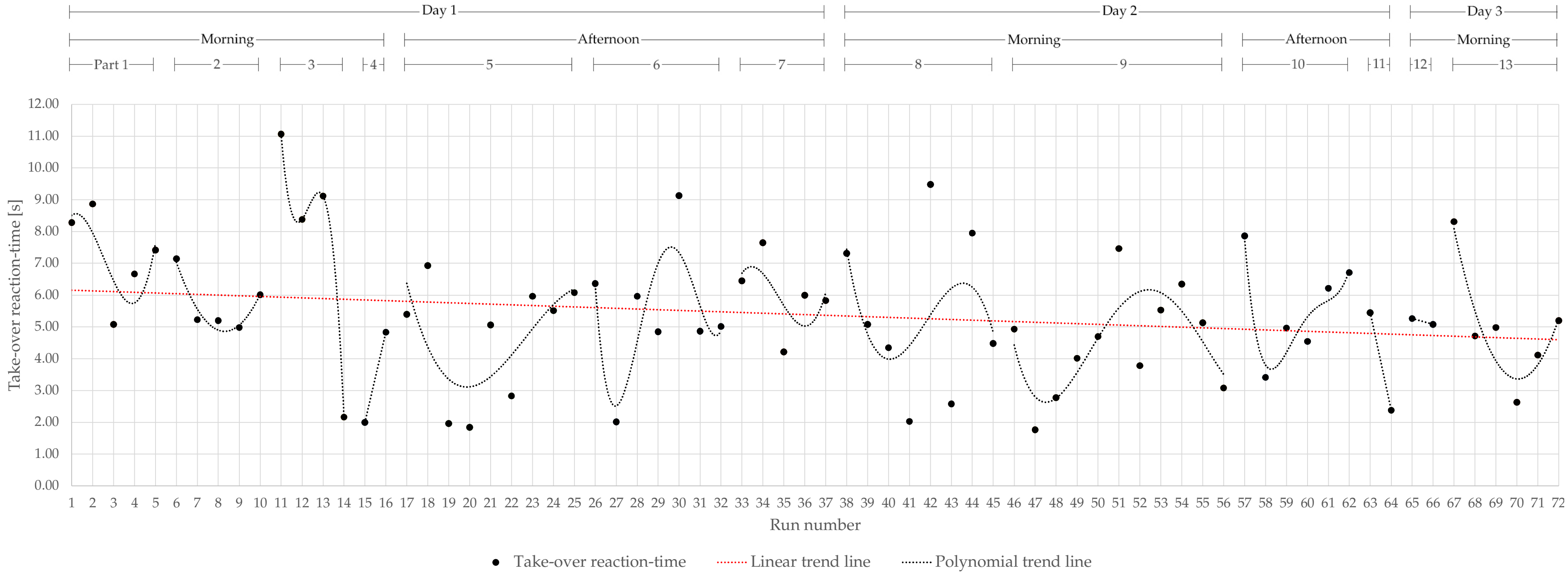


Figure 4.8: Average TOrt of the runs performed on the 1st, 2nd and 3rd day of experimenting.

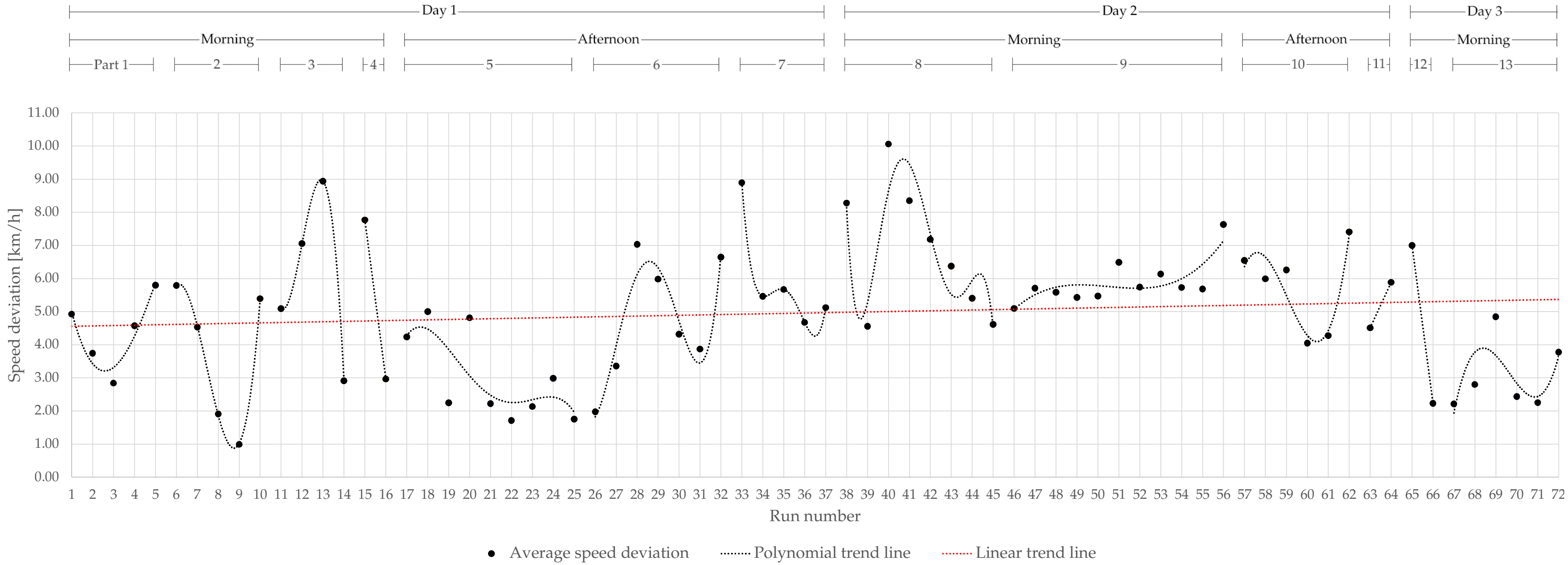


Figure 4.9: Average speed deviation after resumption of manual control after the TOR of the runs performed on the 1st, 2nd and 3rd day of experimenting.

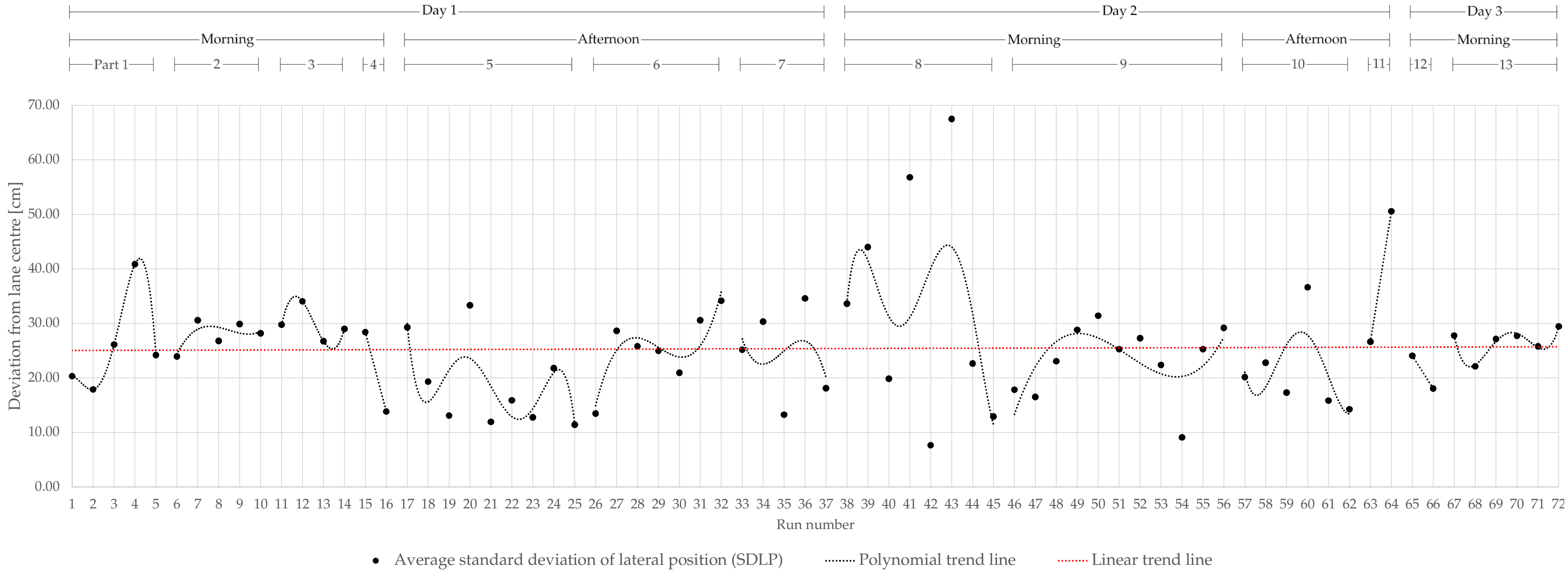


Figure 4.10: Average SDLP after resumption of manual control after the TOR of the runs performed on the 1st, 2nd and 3rd day of experimenting.

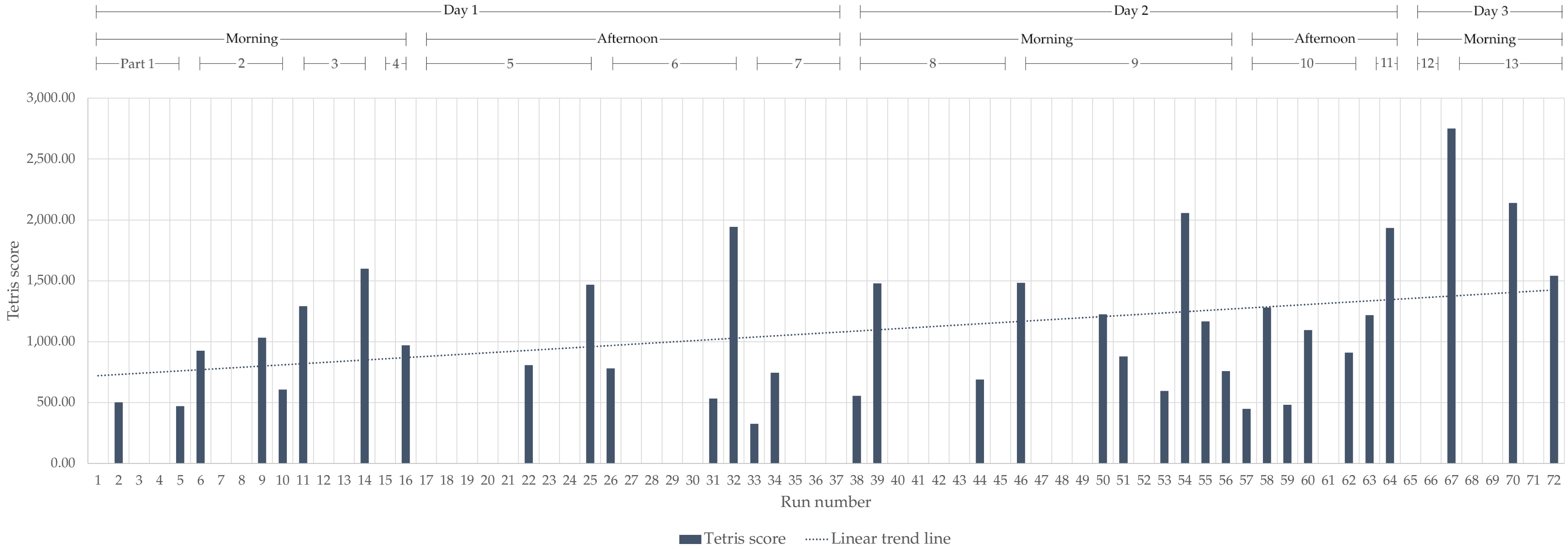


Figure 4.11: Average Tetris score of the runs performed on the three days of experimenting. A Tetris score is only reported for runs with 'game' as NDRT

4.2.5 Main findings

A total of ten scenarios experienced data collection problems, of which for one scenario no data was collected at all and for nine scenarios no finger sensor data is available. Data collection issues with the finger sensor occurred to five other scenarios as well, however, a rerun is available for these scenarios. These reruns will be used for the design variable analysis, instead of the original runs. Thus, the rerun is analysed for subjective and physiological workload as well as for driving and secondary task performance.

In general, few workload trends were found in the measurements of the various workload measures. Trends were found in physiological workload, namely in the mean HR which decreased on average by -0.10 bpm per run and in the effect of the TOR measured by the SDNN which increased on average by +0.17 ms per run. As discussed earlier in Section 3.2.2 and 3.2.9, the physiological workload measures are prone for time-on-task effects. By using an orthogonal design, the effect of time-on-task on the design variables is reduced as much as possible. In addition, it was found that the HF power measurements fell far outside the norm. Thus, the HF power wil not be used as physiological workload indicator in the analysis of the various design variables. Furthermore, a significant trend was found in the Tetris scores, which increased on average by +9.92pt per run. This will also be discussed further in Chapter 5.

The clear workload differences that are measured in the various days, half-days and parts, suggests that, indeed, the design variables affect task demand differently and thus result in different driver workload. Namely, the various days, half-days and parts are not orthogonal, therefore, specific design can appear more or less frequently. Thus, as clear differences in workload are measured, this suggests that the design variables affect driver workload. What design variables does affect driver workload most or least will be analysed in Section 4.3 to 4.6.

4.3 TIME BUDGET AS DESIGN VARIABLE

Four time budgets of 0, 5, 10 and 15 seconds were varied in the experiment. The workload experienced in the scenarios with the different time budgets are compared to analyse the extent to which the different time budgets affected take-over request (TOR)-induced workload. Measures used for analysing driver workload were elaborated in Section 3.2.2. As mentioned in Section 4.2, one scenario was not tested due to an error in the experiment preparation. This scenario had a 0-second time budget, therefore, 15 (instead of 16) scenarios are included in the analysis of 0-second time budgets. The performed statistical analyses is presented in Figure 3.2.8.

4.3.1 Subjective workload

The results of the Raw Task Load Index (RTLX) for the different time budgets are presented in Table 4.11. For every attribute level, the mean and Standard Deviation (SD) of the overall workload and workload on the six RTLX scales are given.

Overall workload

The subjective workload scores for the different time budgets show a decreasing workload trend for every increase in time budget. Especially the 0-second time budget has a large effect on overall workload, with an average score of 46.17%, the other three time budgets result in a workload score of 30.31%, 26.06% and 22.24%, respectively. As the assumption for equal variances did not hold ($F(3, 59) = 7.725$, $p < 0.001$), the Welch and Brown-Forsythe tests were used and showed that the four time budgets were significantly different ($F(3, 31.736) = 4.508$, $p = 0.010$ and $F(3, 40.172) = 6.492$, $p = 0.001$, respectively). The post-hoc Games-Howell test showed that the 0-second time-budget results in a significantly higher workload than the 15-second time budget ($p = 0.008$). All other differences were insignificant.

Workload on the six RTLX scales

With regards to the workload scores on the six scales of the RTLX, a decreasing workload trend is found in the mental, physical, temporal demand and performance scales for every increase in time budget. This trend is not found in the effort and frustration scales. The temporal demand scale reports the greatest difference in workload between the 0-second and 15-second time budget, with a difference of 72.23%. All other workload scales report smaller workload differences between the attribute levels. Only for temporal demand a significant difference is found. The ANOVA showed that the scores for the four time-budgets are significantly different, with $p < 0.001$. The post-hoc Tukey test showed that the 0-second time budget results in a significantly higher temporal demand than all other time budgets ($p < 0.001$ for all three time-budgets), furthermore the 5-second time budget resulted in a significantly higher temporal demand than the 15-second time-budget ($p < 0.001$) and the 10-second time-budget results in a significantly higher temporal demand than the 15-second time budget ($p < 0.001$). Thus, a non-significant difference is only found between the 5- and 10-second time-budget.

Table 4.11: Results of the RTLX for assessing subjective workload in different task load conditions.
MD = Mental Demand; PD = Physical Demand; TD = Temporal Demand; P = Performance; E = Effort; F = Frustration

		Overall [%]	**MD [%]**	**PD [%]**	**TD [%]**	**P [%]**	**E [%]**	**F [%]**
	N(*)	Mean(*SD*)	Mean(*SD*)	Mean(*SD*)	Mean(*SD*)	Mean(*SD*)	Mean(*SD*)	Mean(*SD*)
Average	63(1)	31.29(*17.66*)	33.33(*19.01*)	30.00(*19.20*)	37.70(*29.71*)	20.24(*15.44*)	37.94(*22.01*)	28.57(*20.55*)
Time budget [s]								
0	16(1)	46.17(*22.70*)^{d}	39.33(*25.83*)	40.33(*25.87*)	80.67(*16.89*)^{bcd}	29.67(*18.85*)	47.67(*31.78*)	39.33(*29.39*)
5	16(0)	30.31(*10.96*)	35.00(*13.54*)	28.75(*11.76*)	36.88(*10.94*)^{ad}	20.94(*14.86*)	34.69(*15.54*)	25.63(*15.37*)
10	16(0)	27.39(*14.94*)	32.19(*17.70*)	27.81(*17.41*)	27.50(*17.89*)^{ad}	17.81(*14.94*)	35.63(*16.42*)	23.43(*19.98*)
15	16(0)	22.24(*11.82*)^{a}	27.19(*17.12*)	23.75(*17.37*)	8.44(*8.51*)^{abc}	13.13(*7.72*)	34.38(*20.56*)	26.56(*12.07*)
Traffic density [veh/km/lane]								
0	16(0)	24.69(*18.58*)^{cd}	17.81(*14.72*)	21.56(*19.47*)	34.69(*32.68*)	15.94(*16.75*)	33.75(*23.77*)	24.38(*20.65*)
5	16(0)	34.01(*16.53*)	33.44(*15.35*)	33.13(*17.21*)	39.69(*30.58*)	23.44(*15.46*)	41.56(*20.55*)	32.81(*21.29*)
10	16(0)	33.80(*19.07*)^{a}	38.75(*18.39*)	30.00(*20.90*)	40.94(*30.94*)	23.44(*17.20*)	38.13(*21.82*)	31.56(*21.19*)
15	16(1)	32.78(*16.18*)^{a}	44.00(*17.95*)	35.67(*17.71*)	35.33(*26.56*)	18.00(*11.46*)	38.33(*23.27*)	25.33(*19.59*)
Location take-over request (TOR)								
Straight	32(1)	21.99(*11.85*)^{b}	25.97(*16.35*)^{b}	20.00(*12.45*)^{b}	30.16(*26.09*)	15.16(*15.03*)^{b}	22.74(*13.16*)^{b}	17.90(*15.64*)^{b}
Curved	32(0)	40.31(*17.82*)^{a}	40.47(*18.90*)^{a}	39.69(*19.75*)^{a}	45.00(*31.55*)	25.16(*14.40*)^{a}	52.66(*18.62*)^{a}	38.91(*19.58*)^{a}
Task during automated driving								
Monitoring	32(0)	30.62(*19.47*)	34.53(*22.05*)	29.06(*21.38*)	36.56(*29.61*)	19.06(*15.16*)	37.34(*25.30*)	27.19(*20.48*)
Tetris	32(1)	31.99(*15.88*)	32.10(*15.53*)	30.97(*16.95*)	38.87(*30.27*)	21.45(*15.87*)	38.55(*18.40*)	30.00(*20.86*)

* Number of excluded runs because of data collection errors.
^{abcd} Significant difference at the 1.25% level with first (^{a}), second (^{b}), third (^{c}), fourth (^{d}) attribute level.

4.3.2 Physiological workload

Missing data is reported in some runs that encountered severe noise, for an overview see Table 4.2. The mean Heart Rate (HR), Root Mean Square of Successive Differences (RMSSD) and Standard Deviation of Normal to Normal intervals (SDNN) during the experiment for the different time budgets are presented in Table 4.13a. In Table 4.12 the effect of the take-over on physiological workload is presented by including the increase or decrease in, respectively, HR, RMSSD and SDNN after the take-over. The effect on the High-Frequency (HF)-band is not included in the analysis, as the HF-data is found invalid (Section 4.2.2).

Heart Rate

The mean HR increases for every increase in time budget. However, the ANOVA showed that the differences in mean HR were insignificant: $F(3, 59) = 0.304$, $p = 0.823$. From the table it can also be seen that the TOR, on average, always led to an increase in HR. The effect on HR does not increase with every increase in time budget. Again, the ANOVA showed that the differences were insignificant: $F(3, 50) = 0.768$, $p = 0.517$.

RMSSD (Heart Rate Variability)

The mean RMSSD is almost equal for the 0-, 5-, and 10-second time budget, whereas the 15-second time-budget has the lowest mean RMSSD. The ANOVA showed that the mean RMSSD does not differ for the different time-budgets: $F(3, 50) = 1.040$, $p = 0.124$. Regarding the effect of the take-over on RMSSD, an irregular increase and decrease in RMSSD is found for every increase in time budget. No significant differences are found by performing the ANOVA: $F(3, 50) = 0.284$, $p = 0.837$.

SDNN (Heart Rate Variability)

The highest mean SDNN of 40.46 ms is reported in scenarios with a 0-second time-budget and the lowest SDNN of 31.20 ms is reported for 15-second time budget scenarios. The ANOVA found no significant difference between the time-budgets: $F(3, 50) = 1.566$, $p = 0.209$. Also for SDNN, an irregular increase and decrease is found for every increase in time budget. The 0-, 5-, and 15-second time-budget report, on average, a decrease in SDNN after the take-over, whilst the 10-second time-budget reports an increase in SDNN. The Kruskal Wallis test reports an insignificant difference: $H(3) = 1.746$, $p = 0.627$.

Table 4.12: Physiological workload in the different task load conditions

(a) Mean physiological workload

	HR [bpm]		RMSSD [ms]		SDNN [ms]
	N(*)	Mean(*SD*)	N(*)	Mean(*SD*)	Mean(*SD*)
Average	64(1)	88.19(*7.28*)	64(10)	36.91(*13.95*)	36.87(*12.97*)
Time budget [s]					
0	16(1)	87.64(*9.42*)	16(1)	39.92(*18.13*)	40.46(*15.83*)
5	16(0)	88.49(*5.84*)	16(2)	37.06(*14.67*)	35.80(*9.63*)
10	16(0)	87.14(*5.58*)	16(2)	38.48(*8.08*)	38.57(*8.84*)
15	16(0)	89.47(*8.21*)	16(5)	30.62(*12.11*)	31.20(*8.33*)
Traffic density [veh/km]					
0	16(0)	92.25(*7.91*)	16(2)	29.18(*9.64*)[c]	31.16(*7.45*)[c]
5	16(0)	87.20(*5.47*)	16(3)	36.25(*10.80*)	36.74(*10.47*)
10	16(0)	84.74(*4.93*)	16(3)	48.28(*16.95*)[a]	45.40(*13.62*)[a]
15	16(1)	88.61(*8.69*)	16(1)	34.69(*11.34*)	34.80(*10.00*)
Location take-over request (TOR)					
Straight	32(1)	89.74(*7.95*)	32(4)	34.07(*12.79*)	34.41(*10.05*)
Curved	32(0)	86.69(*6.33*)	32(5)	39.76(*14.72*)	39.34(*12.51*)
Task during automated driving					
Monitoring	32(0)	89.80(*8.50*)	32(5)	34.86(*15.68*)	36.41(*13.26*)
Game	32(1)	86.53(*5.39*)	32(4)	38.97(*11.93*)	37.34(*9.69*)

* Number of excluded runs because of data collection errors.
[abcd] Significant difference at the 1.25% level with first ([a]), second ([b]), third ([c]) fourth ([d]) attribute level.

(b) TOR-induced physiological workload

		HR [bpm]	RMSSD [ms]	SDNN [ms]
	N(*)	Mean(*SD*)	Mean(*SD*)	Mean(*SD*)
Average	64(10)	+ 1.02(*2.72*)	+ 0.31(*11.93*)	- 2.26(*12.97*)
Time budget [s]				
0	16(1)	+ 0.33(*2.87*)	+ 1.20(*14.75*)	- 6.33(*15.47*)
5	16(2)	+ 1.84(*2.88*)	- 0.53(*11.69*)	- 0.49(*12.92*)
10	16(2)	+ 1.11(*2.76*)	+ 2.05(*13.62*)	+ 0.70(*11.10*)
15	16(5)	+ 0.74(*2.27*)	- 2.06(*5.42*)	- 2.73(*11.72*)
Traffic density [veh/km]				
0	16(2)	- 0.26(*2.84*)	- 4.36(*6.86*)	- 2.00(*11.66*)
5	16(3)	+ 1.13(*2.20*)	+ 0.39(*11.56*)	- 2.74(*7.28*)
10	16(3)	+ 2.48(*2.59*)	- 0.53(*18.22*)	- 3.45(*21.90*)
15	16(1)	+ 0.84(*2.72*)	+ 1.73(*10.01*)	- 0.97(*7.40*)
Location take-over request (TOR)				
Straight	32(4)	+ 1.15(*2.45*)	+ 0.73(*8.71*)	- 0.29(*11.09*)
Curved	32(5)	+ 0.89(*3.00*)	- 0.11(*14.63*)	- 4.23(*14.56*)
Task during automated driving				
Monitoring	32(5)	+ 0.52(*2.94*)	+ 0.53(*9.92*)	- 5.82(*14.37*)
Game	32(4)	+ 1.52(*2.42*)	+ 0.09(*13.85*)	+ 1.30(*10.49*)

* Number of excluded runs because of data collection errors.

4.3.3 Driving measures

To analyse driving performance at manual resumption of the Dynamic Driving Task (DDT), the take-over reaction time (TOrt), speed deviation and Standard Deviation of Lateral Position (SDLP) are analysed. The found driving performance metrics for the different time budgets are presented in Table 4.14. The table also presents whether the available time budget was exceeded and how often an exceedance occurred in the different scenarios.

Take-over reaction time (TOrt)

The table shows that the mean TOrt increases with every increase in time budget. A significant difference in TOrts was found by the Kruskal Wallis test: $H(3) = 27.440$, $p < 0.001$. The 0-second time-budget resulted in a significantly lower TOrt compared to all other time budgets, with $p = 0.003$ in comparison to the 5-second time budget, $p = 0.001$ to the 10-second time budget, and $p < 0.001$ to the 15-second time budget.

The time budget was exceeded in 25 scenarios, of which 15 scenarios had a 0-second time budget and 10 scenarios had a 5-second time budget. The duration of the exceeded time budget is significantly longer in scenarios with a 0-second time budget (Kruskal-Wallis $H(1) = 10.351$, $p = 0.001$).

Longitudinal control ability

The speed deviation does not show large differences between the different time-budget scenarios, this is confirmed by the ANOVA that found an insignificant result of $F(3, 59) = 0.050$, $p = 0.985$.

Lateral control ability

The time budget with the greatest effect on lateral control ability of the driver is the 0-second time budget, for which on average a SDLP of 31.00 cm is reported. An ANOVA was performed and found no significant differences: $F(3, 59) = 1.894$, $p = 0.140$.

4.3.4 Secondary task

Table 4.15 presents the average Tetris scores obtained during automated driving before the take-over in the different time budget scenarios. For this analysis, 32 scenarios are analysed as only half the scenarios included playing Tetris as Non-Driving Related Task (NDRT). Higher average scores are reported for scenarios that had a 0- or 5-second time budget compared to scenarios with a 10-, or 15-second time budget. An ANOVA was performed, but found no significant difference, with $p = 0.241$. Thus, the different time budgets do not result in different Tetris scores.

Table 4.14: Analysis of driving performance metrics in different task load conditions

		TOrt [s]	Exceedance [s]		SD speed [km/h]		SDLP [cm]
	N(*)	Mean(*SD*)	N	Mean(*SD*)	N	Mean(*SD*)	Mean(*SD*)
Average	64(1)	5.42(*2.14*)	25	2.27(*1.72*)	64(1)	5.19(*1.93*)	25.59(*10.88*)
Time budget [s]							
0	16(1)	2.96(*1.80*)[bcd]	15	2.96(*1.80*)	16(1)	5.17(*2.02*)	31.00(*15.84*)
5	16(0)	5.66(*1.14*)[a]	10	1.23(*0.90*)	16(0)	5.27(*2.17*)	25.19(*8.35*)
10	16(0)	6.13(*1.70*)[a]	0	-	16(0)	5.27(*1.91*)	22.31(*8.71*)
15	16(0)	6.81(*1.71*)[a]	0	-	16(0)	5.04(*1.79*)	24.19(*8.17*)
Traffic density [veh/km]							
0	16(0)	5.30(*1.90*)	7	3.06(*2.62*)	16(0)	4.12(*2.05*)	16.21(*5.12*)[c]
5	16(0)	4.66(*2.00*)	6	1.90(*0.91*)	16(0)	4.78(*2.28*)	29.92(*14.25*)
10	16(0)	5.39(*2.34*)	6	2.46(*1.53*)	16(0)	5.77(*1.70*)	23.54(*6.90*)[a]
15	16(1)	5.78(*2.61*)	6	1.54(*1.03*)	16(1)	6.12(*1.72*)	32.07(*9.64*)
Location take-over request (TOR)							
Straight	32(1)	5.41(*2.32*)	13	1.62(*1.02*)	32(1)	4.41(*1.83*)[b]	23.87(*11.00*)
Curved	32(0)	5.45(*1.98*)	12	2.98(*1.02*)	32(0)	5.93(*1.75*)[a]	27.25(*10.67*)
Task during automated driving							
Monitoring	32(0)	5.07(*2.27*)	13	1.65(*0.82*)	32(0)	5.18(*2.00*)	25.63(*12.46*)
Game	32(1)	5.80(*1.96*)	12	2.95(*2.18*)	32(1)	5.20(*1.89*)	25.55(*9.18*)

* Number of excluded runs because of data collection errors.
[abcd] Significant difference at the 1.25% level with first ([a]), second ([b]), third ([c]), fourth ([d]) attribute level.

Table 4.15: Analysis of secondary task performance in different task load conditions
* Number of excluded runs because of data collection errors.
[abcd] Significant difference at the 1.25% level with first ([a]), second ([b]), third ([c]), fourth ([d]) attribute level.

		Tetris score
	N(*)	Mean(*SD*)
Average	32(1)	1009.16(*477.50*)
Time budget [s]		
0	8(1)	1158.14(*504.60*)
5	8(0)	1211.88(*549.06*)
10	8(0)	892.75(*346.61*)
15	8(0)	792.50(*446.13*)
Traffic density [veh/km]		
0	8(0)	949.75(*405.88*)
5	8(0)	1140.63(*405.63*)
10	8(0)	1038.25(*378.50*)
15	8(1)	893.57(*338.65*)
Location take-over request (TOR)		
Straight	16(1)	1270.20(*325.12*)[b]
Curve	16(0)	764.44(*298.73*)[a]

4.4 TRAFFIC DENSITY AS DESIGN VARIABLE

Four traffic densities of 0, 5 (low), 10 (medium) and 15 (high) vehicles per kilometre per lane are varied in the experiment. The workload experienced in the scenarios with the different traffic densities are compared to analyse the extent to which the different traffic densities affected take-over request (TOR)-induced workload. Measures used for analysing driver workload were elaborated in Section 3.2.2. As mentioned in Section 4.2, one scenario was not tested due to an error in the experiment preparation. This was a high traffic density scenario with 15 vehicles/km/lane, therefore, 15 (instead of 16) scenarios are included in the analysis of high traffic density. The performed statistical analyses is presented in Figure 3.2.8.

4.4.1 Subjective workload

The results of the Raw Task Load Index (RTLX) for the different traffic densities are presented in Table 4.11. For every attribute level, the mean and Standard Deviation (SD) of the overall workload and workload on the six RTLX scales are given.

Overall workload

Scenarios without traffic resulted, on average, in a lower overall workload score as reported in the table. An ANOVA was performed and found no significant difference: $F(3, 59) = 1.016$, $p = 0.392$.

Workload on the six RTLX scales

With regards to the workload scores on the six scales of RTLX, an increasing workload trend for every increase in traffic density is only found in the mental demand scale. All other scales do not show such an increasing trend. Instead, an irregular increase or decrease in workload has been observed for every increase in traffic density. By performing an ANOVA, a significant difference between the traffic densities is found in the mental demand scale, with $p < 0.001$. A significant difference is found between scenarios without traffic and scenarios with medium and high traffic densities, with $p = 0.004$ and $p < 0.001$, respectively. No significant differences are found for the other RTLX scales. Thus, if no traffic is simulated this results in less mental demand compared to scenarios with 10 or 15 vehicles per kilometre per lane. If traffic is simulated, mild/medium/high traffic densities affect workload to the same extent.

4.4.2 Physiological workload

Missing data is reported in some runs that encountered severe noise, for an overview see Table 4.2. The found Heart Rate (HR), Root Mean Square of Successive Differences (RMSSD) and Standard Deviation of Normal to Normal intervals (SDNN) values before and after the TOR for the different traffic densities are presented in Table 4.13a. Table 4.12 presents the effect of the TOR by including the increase or decrease in, respectively, HR, RMSSD and SDNN after the TOR.

Heart Rate

The mean HR is highest in scenarios without traffic, with an average HR of 92.25. Small differences in mean HR are reported for the other three traffic densities. By performing an ANOVA a non-significant difference is found between the mean HR of the different traffic densities: $F(3, 50) = 3.311$, $p = 0.026$ (Bonferroni correction). The mean HR does not differ for the different traffic densities. With regards to the effect of the take-over on HR, an increasing effect is found for every increase in traffic density, except for the high traffic density condition that reports a smaller effect on HR compared to the medium traffic density condition. However, a non-significant difference between the effect on HR is found by performing the ANOVA: $F(3, 50) = 2.518$, $p = 0.069$.

RMSSD (Heart Rate Variability)

The mean RMSSD increases for every increase in traffic density, except for the high traffic density scenarios. The ANOVA was performed and found a significant difference between the mean RMSSD for the different traffic densities: $F(3,50) = 5.589$, $p = 0.002$. A significant difference of $p = 0.001$ is found between zero and medium traffic density scenarios. Thus, the low and high traffic density scenarios do not result in a different RMSSD than zero or medium traffic densities, but the medium traffic density scenarios result in a higher RMSSD than scenarios without traffic.

As regards to the effect of the take-over on RMSSD in the different scenarios, an irregular increase and decrease is found. The ANOVA was performed and found a non-significant difference of the effect of traffic densities on RMSSD: $F(3,50) = 0.099$, $p = 0.960$.

SDNN (Heart Rate Variability)

The mean SDNN shows a similar pattern as the RMSSD for the different traffic densities. An ANOVA was performed and found a significant effect: $F(3,50) = 5.589$, $p = 0.008$. Likewise the RMSSD, a significant difference is found between scenarios with a zero and medium traffic density, with $p = 0.005$.

The effect of the take-over on SDNN is not similar to the RMSSD. Instead, for every traffic density it reports a decrease in SDNN after the take-over. The effect increases for every increase in traffic density, except for the high traffic density scenarios that report the smallest average effect on SDNN. A Kruskal Wallis test was performed and found, however, no significant differences between the traffic densities on the effect on SDNN: $H(3) = 0.598$,$p = 0.897$.

4.4.3 Driving measures

To analyse driving performance at manual resumption of the Dynamic Driving Task (DDT), the take-over reaction time (TOrt), speed deviation and Standard Deviation of Lateral Position (SDLP) are analysed. The TOrts found for the different traffic densities are presented inTable 4.14. The table also presents whether the available time budget was exceeded and how often an exceedance occurred in the different scenarios.

Take-over reaction time

The table presents an irregular TOrt for every increase in traffic density. The shortest TOrt is reported for scenarios with low traffic density, whereas the longest TOrt is reported for scenarios with high traffic density. A Kruskal Wallis test was performed and, indeed, found no significant difference in TOrt for the different traffic densities: $H(3) = 0.873$, $p = 0.832$.

The distribution of scenarios in which the time budget was exceeded shows a constant pattern, it is exceeded 7 times in scenarios with zero traffic density and is exceeded 6 times in scenarios with a low, medium, and high traffic density. The duration of exceeded time budget does also not show a regular pattern. The longest exceedance is reported in scenarios without any traffic with an average exceedance of 3.06 seconds. Whereas the shortest average exceedance is reported in scenarios with a low traffic density, with an average exceedance of 1.90 seconds. The differences in duration of the exceedance is found non-significant by performing a Kruskal Wallis test: $H(3) = 2.241$, $p = 0.524$.

Longitudinal control ability

With regards to the average speed deviation at different traffic densities, an increasing speed deviation is found for every increase in traffic density. However, the ANOVA found a non-significant difference between the traffic densities: $F(3,50) = 3.201$, $p = 0.030$ (Bonferroni correction).

Lateral control ability

The average SDLP for the different traffic densities does show an increase for every increase in traffic density, except for the medium traffic density scenarios. An ANOVA was performed and found a significant difference between the average SDLP of the different traffic densities: $F(3, 62) = 8.193$, $p < 0.001$. The SDLP in scenarios without traffic was significantly smaller compared to the SDLP found in low ($p = 0.007$), medium ($p = 0.002$), and high ($p < 0.001$) traffic density.

4.4.4 Secondary task

The Tetris scores obtained during automated driving in the different traffic density scenarios are presented in Table 4.15. In scenarios with low and medium traffic density higher Tetris scores were obtained than in scenarios without traffic or with a high traffic density. An ANOVA was performed, and found no significant differences in Tetris scores for the different traffic densities, with $p = 0.959$.

4.5 LOCATION OF THE TAKE-OVER AS DESIGN VARIABLE

Two locations of the take-over are varied in the experiment: a take-over on a straight road section and in a curve. The workload experienced in the scenarios with the different locations are compared to analyse the extent to which the different take-over request (TOR) locations affected TOR-induced workload. Measures used for analysing driver workload were elaborated in Section 3.2.2. As mentioned in Section 4.2, one scenario was not tested due to an error in the experiment preparation. In this scenario the TOR was issued on a straight road section, therefore, 31 (instead of 32) scenarios are included in the analysis of TORs on straight road section. The performed statistical analyses is presented in Figure 3.2.8.

4.5.1 Subjective workload

The results of the Raw Task Load Index (RTLX) for the different TOR locations are presented in Table 4.11. For every attribute level, the mean and Standard Deviation (SD) of the overall workload and workload on the six RTLX scales are given.

Overall workload

Scenarios with a take-over in a curve resulted in a hibeats per minute (bpm)her subjective overall workload score compared to the take-over in a straight road section, with a score of 40.31 and 21.99, respectively. The Welch and Brown-Forsythe tests showed that this workload difference is significant: $F(1, 54.100) = 23.228$, $p < 0.001$ in both tests.

Workload on the six RTLX scales

All six RTLX scales report a higher score for scenarios with a take-over in a curve. Only the temporal demand scale was found non-significant by the ANOVA. An ANOVA found a significant difference for the mental demand scale ($p = 0.002$), performance scale($p = 0.009$), and frustration ($p < 0.001$). The welch and Brown-Forsythe tests found a significant difference for the physical demand scale ($p < 0.001$ in both tests) and effort scale ($p < 0.001$ in both tests).

4.5.2 Physiological workload

Missing data is reported in some runs that encountered severe noise, for an overview see Table 4.2. The found Heart Rate (HR), Root Mean Square of Successive Differences (RMSSD) and Standard Deviation of Normal to Normal intervals (SDNN) values before and after the TOR for the different TOR locations are presented in Table 4.13a. Table 4.12 presents the effect of the TOR by including the increase or decrease in, respectively, HR, RMSSD and SDNN after the TOR.

Heart Rate

The mean HR in scenarios with a take-over on a straight road section was 89.74 HR and for curve scenarios it was 86.69 HR on average. This difference is found insignificant by the ANOVA: $F(1, 61) = 2.850$, $p = 0.096$. With regards to the effect on HR of the take-over in the different scenarios, both scenarios report on average an increase in HR after the take-over. Again, this difference was found non-significant by the ANOVA: $F(1, 52) = 0.124$, $p = 0.726$.

RMSSD (Heart Rate Variability)

Scenarios with a take-over in a curve report, on average, a higher RMSSD than scenarios with a take-over in a straight road section: 39.76 and 34.07 ms, respectively. This difference is found non-significant by the ANOVA: $F(1, 52) = 2.301$, $p = 0.135$. With regards to the effect on RMSSD of the take-over, an increase of 0.73 ms is reported in straight road section scenarios, whereas a decrease of 0.11 ms is reported in curved scenarios. The Welch and Brown-Forsythe tests found that this difference is non-significant: $F(1, 42.394 = 0.065$, $p = 0.799$ in both tests.

SDNN (Heart Rate Variability)

Likewise the RMSSD, the mean SDNN is higher in scenarios with a take-over in a curve (39.34 ms) compared to scenarios with a take-over on a straight road section (34.41). However, this difference is found non-significant by the ANOVA: $F(1, 52) = 2.542$, $p = 0.117$. It is found that the take-over results in a decrease in SDNN for both scenarios: -0.29 ms in straight road scenarios and -4.23 in curved scenarios. However, again this difference is found non-significant by the Kruskal Wallis test: $H(1) = 0.399$, $p = 0.528$.

4.5.3 Driving measures

To analyse driving performance at manual resumption of the Dynamic Driving Task (DDT), the take-over reaction time (TOrt), speed deviation and Standard Deviation of Lateral Position (SDLP) are analysed. The driving performance metrics found for the different take-over locations are presented in Table 4.14. The table also presents whether the available time budget was exceeded and how often an exceedance occurred in the different scenarios.

Take-over reaction time

An almost equal TOrt for the two take-over locations is found, with 5.41 seconds on average for take-overs on a straight road section and 5.45 seconds for take-overs in a curve. As expected, this difference is found to be non-significant by the Kruskal Wallis test: $H(1) = 0.048$ $p = 0.826$.

The distribution of exceeded time budgets is almost equal between the two locations, with 13 exceedances in scenarios with a TOR in a straight road section, and 12 exceedances in scenarios with a TOR in a curve. The duration of the exceedance, however, shows a clear difference between the two locations, with an average exceedance of 2.98 seconds for scenarios in which the TOR was issued in a curve, and 1.85 seconds for scenarios in which the TOR was issued on a straight road section. The difference in exceedance between the two TOR locations was found not to differ by performing a Kruskal Wallis test: $H(1) = 3.834$, $p = 0.050$.

Longitudinal control ability

With regards to the average speed deviation, when the take-over occurred in a straight road section a speed deviation of 4.41 km/h is found, whereas a higher speed deviation of 5.93 km/h is found in scenarios with a take-over in a curve. This is found significant by the ANOVA: $F(1, 61) = 11.394$, $p = 0.001$.

Lateral control ability

The average SDLP does not show a large difference between the two take-over locations, with 23.87 cm in scenarios with a take-over in a straight road section and 27.25 cm in a curve. This difference is found non-significant by the ANOVA: $F(1, 61) = 1.531$, $p = 0.221$.

4.5.4 Secondary task

For secondary task performance a large difference in Tetris scores is found: 1,270.20 for straight road sections and 764.33 for curves. This difference is found significant by the ANOVA, with $p = 0.002$.

4.6 NON-DRIVING RELATED TASK AS DESIGN VARIABLE

Two tasks during automated driving are varied in the experiment: a monitoring task and playing Tetris as Non-Driving Related Task (NDRT). The workload experienced in the scenarios with the different tasks are compared to analyse the extent to which the different tasks affected take-over request (TOR)-induced workload. Measures used for analysing driver workload were elaborated in Section 3.2.2. As mentioned in Section 4.2, one scenario was not tested due to an error in the experiment preparation. In this scenario, the participant had to play Tetris as NDRT during automated driving. Therefore, 31 (instead of 32) scenarios are included in the analysis of Tetris as task during automated driving. The performed statistical analyses is presented in Figure 3.2.8.

4.6.1 Subjective workload

The results of the Raw Task Load Index (RTLX) for monitoring and playing Tetris are presented in Table 4.11. For both tasks, the mean and Standard Deviation (SD) of the overall workload and the six scales are given.

Overall workload

With regards to overall workload, Tetris on average resulted in a higher workload than the monitoring task, with a overall workload of 30.62% and 31.99%, respectively. The ANOVA was performed and found no significant difference: $F(1, 61) = 0.093$, $p = 0.762$.

Workload on the six RTLX scales

With regards to the workload scores on the six scales of the RTLX, small differences are found between the two tasks. ANOVAs and Welch-Brown Forsythe tests were performed and found no significant differences between the tasks.

4.6.2 Physiological workload

Missing data is reported in some runs that encountered severe noise, for an overview see Table 4.2. The found Heart Rate (HR), Root Mean Square of Successive Differences (RMSSD) and Standard Deviation of Normal to Normal intervals (SDNN) values before and after the TOR for the different tasks are presented in Table 4.13a. Table 4.12 presents the effect of the TOR by including the increase or decrease in, respectively, HR, RMSSD and SDNN after the TOR.

Heart Rate

Scenarios with a monitoring task, on average, report a higher average HR of 89.80 compared to the Tetris task with a HR of 86.53 on average. This difference is found non-significant by the ANOVA: $F(1, 61) = 3.287$, $p = 0.075$.

With regard to the effect on HR, playing Tetris as game resulted in the greatest increase in HR, with an average increase of 1.47 HR. Also the monitoring task resulted in an increase in HR, with an average increase of 0.52. This difference is found non-significant by the ANOVA: $F(1, 52) = 1.880$, $p = 0.176$.

RMSSD (Heart Rate Variability)

The mean RMSSD in scenarios with a monitoring task was 34.86 ms and was 38.97 ms in Tetris scenarios. An ANOVA was performed, but found a non-significant difference: $F(1, 52) = 1.173$, $p = 0.284$. With regards to the effect of the take-over on RMSSD, for both scenarios an increase in RMSSD is found (+0.53 ms and +0.09, respectively). However, again, this difference is found to be non-significant by performing an ANOVA: $F(1, 52) = 0.892$, $p = 0.892$.

SDNN (Heart Rate Variability)

The mean SDNN between the scenarios reports a small difference, with 36.41 ms for monitoring and 37.34 for Tetris. The ANOVA was performed and found no significant difference: $F(1, 52) = 0.086$, $p = 0.771$. As regards to the effect on SDNN of the take-over a decrease of 5.82 ms is found on average for monitoring scenarios and an increase of 1.30 ms is found for Tetris scenarios. However, this difference is found non-significant by the Kruskal Wallis test: $H(1) = 3.023$, $p = 0.082$.

4.6.3 Driving measures

To analyse driving performance at manual resumption of the Dynamic Driving Task (DDT), the take-over reaction time (TOrt), speed deviation and Standard Deviation of Lateral Position (SDLP) are analysed. The driving performance metrics found for the different tasks are presented in Table 4.14. The table also presents whether the available time budget was exceeded and how often an exceedance occurred in the different scenarios.

Take-over reaction time

A larger mean TOrt is found for scenarios in which Tetris was played during automated driving, with an average TOrt of 5.80 seconds. The mean TOrt of monitoring scenarios is 5.07 seconds. A Kruskal Wallis test was performed and found no significant difference between the TOrts: $H(1) = 2.084$, $p = 0.149$.

The time budget was exceeded 13 times in scenarios when the driver was monitoring the system and was exceeded 12 times when Tetris was played. Tetris scenarios report an average higher exceedance of 2.95 seconds, whereas monitoring scenarios report an average exceeedance of 1.65 seconds. The duration of the exceedance was significantly higher with Tetris as NDRT (Kruskal Wallis: $H(1) = 4.734$, $p = 0.030$).

Longitudinal control ability

With regards to the average speed deviation of the different scenarios, a small difference is reported with a speed deviation of 5.18 km/h in monitoring scenarios and 5.20 km/h in Tetris scenarios. As expected, the ANOVA found non-significant differences: $F(1,61) = 0.002$, $p = 0.968$.

Lateral control ability

The mean SDLP also does show a small difference between the scenarios, with 25.63 cm on average in monitoring scenarios and 25.55 cm in Tetris scenarios. As expected, the ANOVA found no significant differences: $F(1,61) = 0.001$, $p = 0.978$.

5 DISCUSSION

The results of the study are discussed in this chapter. First, the participant selection and analysis methodology for the personality experiment are discussed in Section 5.1. Subsequently, the results of the self-experiment are discussed. First an overview is given of the average effect of a take-over on the driver workload Section 5.2. Then in Section 5.3 the role of the different designs on driver workload is discussed. After that, in Section 5.4 the used driver workload measures are discussed. Finally, in Section 5.5 recommendations were made for the future research on personality or other driving simulation studies, regarding the design of the scenario, regarding the apparatus and regarding the analysis of driver workload.

5.1 PERSONALITY EXPERIMENT

Before COVID-19, registration for participating in the driving simulation experiment regarding personality was already open and 159 people registered. A selection method was presented that assigned an applicant to the personality trait for which the highest score was obtained and subsequently selected the 20 highest scoring individuals per trait. The relatively most extreme individuals are selected, which explains the low mean score of the neuroticism group of 28.15 out of 40 points, while all other personality traits report a mean closer to the maximum, see Table 4.1. Ideally, individuals would be selected who achieved a (nearly) maximum score on one of the personality traits and a below-average score on the other. However, a person's personality is a mixture of these traits. For example, one individual who applied for the experiment achieved the maximum of 40 points for extraversion, 44 out of 45 points for agreeableness, and 49 out of 50 points for openness. This individual was selected for extraversion and would be analysed for this specific trait.

This immediately shows the downside of using the aforementioned analysis method, namely driver behaviour of this specific person is related to the combination of extraversion, agreeableness and openness. Analysing persons only on their dominant trait, ignores the interaction between the traits. For example, Taubman-Ben-Ari and Yehiel (2012) provided a classification of driving styles based on a combination of Big Five personality traits. For example, the reckless and angry driving style is characterised by high levels of extraversion and lower levels for agreeableness and conscientiousness. Therefore, it it recommended to analyse the participants based on their set of personality traits. This recommendation is not an argument in favour of using the Big One instead of the Big Five, as the Big Five is a validated method for distinguishing personalities. Instead, it is a recommendation for using the Big Five to its fullest potential.

5.2 REFLECTION ON THE VALIDATION EXPERIMENT

Although design variables and respective attributes were chosen, which were expected to result in significant workload differences, this was only found on a limited scale. This has two reasons. Firstly, because of the $N = 1$ study design, the number of observations was limited. Secondly, the fact that the number of variables to be analysed in this study was maximised in as few runs as possible, resulting in large Standard Deviations (SDs) in the workload measurements. It is therefore already exceptional that significant differences were found. The workload differences found between the various designs provide sufficient indication of finding significant workload differences in the future experiment regarding personality (with >100 participants). This indication is reinforced when significant differences were demonstrated in the current $N = 1$ study.

5.2.1 Take-over requests and their effect on workload

All scenarios combined, an overall workload as measured by the Raw Task Load Index (RTLX) of 31.29% was found, which lies within the expected range for subjective workload measured by the RTLX in simulated automated driving studies (Heikoop et al., 2019a). However, many studies only require the participant to respond once to a take-over request (TOR), or use a long duration in between TORs. Therefore, it is important to compare the found overall workload score to a study that also required the participants to regularly respond to TORs in short time span. Namely, in this study, on average, every five minutes a new experiment run was started in a series of consecutive runs. In between runs a short time was used for setting up the next experiment run and to check correct placement of the finger sensor and the chest strap. The study by De Winter et al. (2016) is similar to the current study as it required the participants to respond to TORs every three minutes. For the two experiment that were part of the study by De Winter et al. (2016), an average overall workload of 31% was measured, which is comparable to the finding in the current study. However, the reported workload scores on the six RTLX scales differ from the scores found by De Winter et al. (2016). For instance, this study reports an average frustration of 28.96%, compared to 45% and 46% in De Winter et al. (2016). Higher frustration in this study was expected because of the design of the experiment that required the participant to either monitor the Automated Driving System (ADS) and vehicle or to play Tetris during automated driving. Monitoring requires sustained attention to the task in order to respond timely to the TOR, which is related to increased frustration Warm et al. (2008). Less frustration is expected when the driver is engaged in a Non-Driving Related Task (NDRT) during automated driving. However, the short duration of the NDRT before the TOR was also expected to result in increased frustration, as it was expected to lower the willingness to take over control Hock et al. (2018).

For physiological workload, a mean Heart Rate (HR) of 88.19 beats per minute (bpm) was measured during the experiment, which lies within one SD of the mean resting HR of M(*SD*) = 80.2(14.8) (specifically, for people aged between 21 and 30 years, like the researcher herself) (Avram et al., 2019). The measured Root Mean Square of Successive Differences (RMSSD) of 36.91 ms and Standard Deviation of Normal to Normal intervals (SDNN) of 36.87 ms fall within the standards for short-term measurements of M(*SD*) = 42(15) ms for RMSSD and M(*SD*) = 50(16) ms for SDNN (Shaffer and Ginsberg, 2017). Simulated automated driving was therefore not experienced as very demanding, but it could also not be considered an easy or relaxing activity. The average increase of $+1.02$ bpm in HR and decrease of -2.26 ms in SDNN after the TOR indicate that the TOR increased workload. In contrast, the average increase of $+0.31$ ms in RMSSD after the TOR indicates a reduced workload after the TOR. Alrefaie et al. (2019) and Ruscio et al. (2017) also measured the difference in HR after a TOR and reported an increase in HR of, respectively, $+0.43$ bpm and $+2.98$ bpm after the TOR. The results of this study are thus within the expected range, although based on Ruscio et al. (2017), it could be argued that these results indicate a small effect size.

This study found a relatively high take-over reaction time (TOrt) of 5.42 seconds, which is considerably longer than the TOrt found by Gold et al. (2013) of 2.06 and 3.10 seconds for 5- and 7-second time budgets, respectively. The used design of the take-over attributed to this increased TOrt. Usually, a button on the steering wheel or pressing a driving pedal is used to resume manual driving before the expiration of the time budget. However, this study required the participant to press a key combination on the keyboard to resume driving before the expiration of the available time budget. As the keyboard was placed on the dashboard of the simulator, the driver had to reposition to be able to press the key-combination (Section 3.2.7). Therefore, a longer TOrt is found in this study compared to other studies. The average TOrt which was found in scenarios with a 0-second time budget of 2.96 seconds (i.e. when no key combination is needed to resume control), is comparable to the TOrt found by Gold et al. (2013). The other driving performance measures reported values within the expected range, with an average speed variation of 5.19 km/h and Standard Deviation of Lateral Position (SDLP) of 25.59 cm. For example, Reimer et al. (2007) found a speed deviation of 5.9 km/h, and Naujoks et al. (2014) reported an average SDLP between 15 and 30 cm after a take-over at a speed of 50 km/h.

5.3 DESIGN VARIABLES AND THEIR EFFECT ON WORKLOAD

Time budget

Increases in time budget were expected to decrease take-over request (TOR)-induced workload (Gold et al., 2013). The Raw Task Load Index (RTLX) measurements were in accordance with the aforementioned hypothesis. Namely, an overall workload as measured by the RTLX of 46.17% was found for scenarios with a 0-second time budget and 22.24% for scenarios with a 15-second time budget. However, only the difference in workload between these two time budgets was found significant. Nearly all RTLX scales showed a decrease in workload for every increase in time budget. Especially so (and evidently) for the temporal demand scale, which showed large workload differences between the time budgets. Only workload between the 5- and 10-second time-budget was found non-significant. Based on Gold et al. (2013), a larger difference in perceived workload between 5- and 10-second time budgets was expected. Namely, already with a 7-second time budget, participants in the study by Gold et al. (2013) were able to significantly improve their take-over behaviour as drivers were able to analyse the driving environment before taking over. For 5-second time budgets, drivers did not use the available time to analyse the driving environment. Therefore, it was expected that the 5-second time-budget would result in a higher overall workload compared to the 10-second time budget in the current study.

As for physiological workload, the measurements did not indicate an increase in workload with every increase in time budget. Heart Rate (HR) suggested that the 5-second time budget had the greatest effect on workload, whereas the Root Mean Square of Successive Differences (RMSSD) indicated that the 15-second time budget has the greatest effect on workload. Only Standard Deviation of Normal to Normal intervals (SDNN) indicated that the 0-second time budget had the greatest effect on workload, as was expected. In contrast, SDNN and RMSSD indicate that the 10-second time budget had the least effect on workload (in fact, suggests that workload was lower after the TOR). HR indicates that the 0-second time budget had the least effect on workload. All in all, the physiological workload measurements did not appear to correlate with the RTLX workload measurements.

The take-over reaction time (TOrt) increased with every increase in time budget, despite the intention to take over as soon as possible after the TOR regardless of the duration of the time budget. This was expected, as a longer time budget reduces the rush to take over as soon as possible (Gold et al., 2013). The TOrt in 0-second scenarios was found to be significantly faster than of scenarios with a 5-, 10-, or 15-second time-budget. However, the differences in TOrt between the 5-, 10-, or 15-second time budget were not significantly different. The speed deviation shows little differences between the time budgets. The Standard Deviation of Lateral Position (SDLP), although not significant, does suggest that a 0-second time-budget leads to more workload compared to other time budgets. However, SDLP in scenarios with a 5-, 10-, and 15-second time budgets was almost equal.

As expected, there was no significant difference in secondary task performance between scenarios differing in the duration of the time budget. Nor do the average reported scores suggest the existence of a relation between the Tetris score and the duration of the time budget.

Summarising, the various workload measures did not unambiguously indicate differences in TOR-induced workload with different time budgets. It was expected that every increase in time budget would result in a smaller effect of the TOR on workload. The RTLX suggests, indeed, that workload decreases with a longer available duration to respond to the TOR. Similarly, the SDNN and SDLP suggests that the 0-second time budget has a greater effect on workload than time budgets of 5 seconds or greater. However, driving performance, as indicated by the SDLP, did not seem to improve with a time budget duration greater than 5 seconds. Moreover, the SDNN provides no evidence for workload differences between time budgets of 5 seconds or greater. Thus, although the RTLX was sensitive to variations in task demand, the physiological workload measurements and driving performance seemed less sensitive; only sensitive to high task demands. So if a high demand take-over is to be simulated, it is recommended to use a 0-second time budget instead of a 5-second time budget. Moreover, time budgets of 15 seconds are recommended for low demand take-overs, as the RTLX found a significant difference between 0- and 15-second time budgets. Although, it could also be argued that a 10-second time budget could be used for non-urgent take-overs. Namely, already in this $N = 1$ study a significance of $p = 0.056$ was found for the difference between the 0- and 10-second time budget.

Traffic density

Lower traffic densities were expected to lead to less workload (Gold et al., 2018). The RTLX measurements were only partly in accordance to this hypotheses. Namely, scenarios without traffic resulted in a low overall workload of 24.69%, whereas low traffic density reported an overall workload of 34.01%, 33.80% was reported for medium traffic density, and 32.78% for high traffic density. A significant difference was only found between zero, medium and high traffic density, the difference with the low traffic density was found to be non-significant. As for the six RTLX scales, only mental demand shows an increase in workload for every increase in traffic density. It is interesting to note that temporal demand, performance and frustration show an increase in workload from zero to low traffic density, after which workload remained constant between low and medium traffic density, and then decreased from medium to high traffic density. So, the RTLX appeared sensitive in distinguishing task demands for the various time budget, but appears less sensitive in distinguished between the various traffic densities. This suggests that low, medium and high traffic density at the take-over location did not result in different task demands.

Regarding physiological workload, HR decreased after the TOR in no-traffic scenarios and increased in scenarios with low, medium and high traffic density. Contrary to expectations, take-overs in medium traffic density resulted in a greater increase in HR than take-overs at low and high traffic density. In contrast to the HR, RMSSD reports the largest increase in workload after the TOR for no-traffic scenarios. Furthermore, scenarios with medium traffic density report an increase in RMSSD, whereas it reports a decrease in workload for low and high traffic density. This raises the question if RMSSD is an appropriate workload measure in this experiment. Namely, the RMSSD did report a different direction of the effect on workload than was expected, for both traffic density as time budget as discussed earlier. As regards to the SDNN, the measurements all show an increase in workload after the TOR for all traffic densities. The greatest increase is reported for medium traffic density scenarios and the least increase for high traffic density. Thus, both HR and SDNN appeared sensitive to variations in traffic density. However, take-overs in high traffic density did not appear to have resulted in the greatest effect on physiological workload, even though this was expected based on literature (Radlmayr et al., 2014).

It was expected that the TOrt would increase with every increase in traffic density, as was found in Gold et al. (2016). The results of this study, indeed, suggests that TOrt increases with increases in traffic density. However, in scenarios without any traffic, a similar TOrt was found as in scenarios with medium traffic density (i.e. a longer TOrt as compared to the low traffic density scenarios). This suggests that, indeed, when traffic density increases, workload increases as well. However, a longer TOrt does not necessarily indicate increased driver workload. Namely, without traffic, it does not matter how long it takes for the driver to take over, as no one is bothered.

Speed deviation, although not significant, does suggest an increase in workload with increases in traffic density, from a speed deviation of 4.12 km/h in no-traffic scenarios to a Standard Deviation (SD) of 6.12 km/h in high traffic densities. Increases in speed deviation are not due to overtaking manoeuvres as the speed deviation around an overtaking is excluded in the calculation. Regarding the SDLP, the smallest SDLP of 16.21 cm is found in no-traffic scenarios and the highest SDLP of 32.07 cm is found in high traffic density scenarios. However, medium traffic density reported the second smallest SDLP of 23.54 cm, whilst in low traffic density scenarios an average SDLP of 29.92 cm was found, which is similar to the high traffic density scenarios. SDLP seemed sensitive to distinguish low from high task demand, although appeared less sensitive in distinguishing between medium and high task demands.

As mentioned in Section 3.2.9, traffic density of the scenario (thus, also at the take-over location) was already revealed to the driver at the start of the experiment. Therefore, this could have affected the Tetris scores, as possibly the driver would be less engaged in playing Tetris at high demand scenarios. However, there was no significant difference in secondary task performance between scenarios differing in traffic density. Neither do the average reported scores suggest the existence of a relation between the Tetris score and the traffic density. Thus, task engagement did not differ between scenarios differing in traffic density. However, as traffic density is already revealed to the driver at the start of the experiment, this could have prevented the finding of larger workload differences.

In summary, it has been found that driving performance after the TOR decreased with increases in traffic density in the current experimental set-up. Possibly, overload developed due to increases in traffic density, which is detrimental for driving performance (Endsley, 2019). However, the subjective and physiological workload measures did not fully indicate overload in high traffic density scenarios. Namely, less workload was measured in scenarios with high traffic density by the RTLX, HR, and SDNN. So, if in a future experiment, a low demand take-over is simulated, it is preferred to use zero traffic density over a low traffic density of 5 vehicles/km/lane, as the low and medium traffic density induced similar workload. If high task demand must be varied, it is recommended to simulate a low or medium traffic density of 5 or 10 vehicles per kilometre per lane. However, preference is given to simulate a medium traffic density, as already in this experimental set-up, significant workload differences were found between no-traffic and medium traffic density scenarios.

Location of the take-over

Based on Mok et al. (2015) and Naujoks et al. (2017), it was hypothesized that a take-over in a curve would result in a significantly more TOR-induced workload than a take-over on a straight road section. The RTLX subjective workload measurements were in accordance to the aforementioned hypothesis, as a significant workload difference was found between workload induced by a TOR in a curve or a straight road section. The TOR in a straight road section resulted in a low mean overall workload of 21.99%, whereas the TOR in a curve resulted in an almost double overall workload of 40.31%. All RTLX scales, expect temporal demand, reported a significant workload difference between the locations. Even though temporal demand was non-significant, it does suggest that a TOR in a curve induces more workload than a TOR in a curve, with a mean temporal demand of 30.16% for the TOR on a straight road section and 45.00% for a TOR in a curve. A difference in temporal demand was expected, as a timely response to the TOR in a curve is critical to keep to the driving lane.

Contrary to expectations, the HR shows a greater effect of the take-over in scenarios with a TOR on a straight road section. The HR graph in Figure 5.2c shows a small difference in the effect on HR. Scenarios with a TOR on a straight road section report two peaks shortly after the TOR, whereas the HR in curve scenarios increases immediately after the TOR and remains elevated for approximately 10 seconds until it returns to the HR level of before the take-over. The RMSSD reports a decrease in workload when the TOR in issued on a straight road section, and reports an increase when it is issued in a curve. This is the first time that the RMSSD indicates a direction of workload that was expected based on literature. The SDNN, again as expected, indicated that a TOR in a curve increases workload more than a TOR on a straight road section. In fact, a small decrease in workload in found after a TOR that is issued on a straight road section.

The found mean TOrts are almost identical. However, a difference is found in the duration of the exceedance of the time budget, which is greater for TORs in a curve. This is counter-intuitive, as explained earlier. A reason could be found in the method of TOrt calculation, which measures the time between the TOR and the first acceleration. Namely, the driver might not press the gas pedal to reach a lower speed which is preferable when driving in a curve. However, the ego-vehicle is always trying to maintain a constant speed of 120 km/h during automated driving, also when driving in a curve. Thus, the vehicle could be at a higher speed than preferred at the take-over. The driver, therefore, delays acceleration after the take-over to reach the desired speed for driving in a curve. Regarding the other performance measures, a significantly greater speed deviation after the take-over is found in scenarios with a TOR in a curve. Also, the SDLP is greater when the TOR is issued in a curve.

As expected, there was a significant difference in secondary task performance between scenarios with a different TOR location. As explained before, the duration of manual driving of both scenarios is not equal, thus differences in Tetris score were expected. Namely, it was expected that an increased game duration, would increase the average Tetris score. Indeed, straight scenarios that are located further in the scenario, report a higher average Tetris score.

In summary, in the current experiment set-up, the TOR in a curve induced more workload compared to a TOR on a straight road section. The difference in workload was most evident in the difference in subjective workload measured by the RTLX. Physiologically, the difference in workload is also found by the RMSSD and SDNN. Post-take-over driving performance is noticeably worse when the TOR was issued in a curve, as measured by the speed deviation and SDLP. However, a distorted TOrt is possibly found; the mean TOrt in scenarios with a TOR in a curve is possibly shorter than measured. The used TOrt calculation method could have attributed to the increased TOrt, which measured the time between the TOR and the first acceleration. As the speed of the vehicle at the take-over in a curve is approximately 120 km/h, the driver delays acceleration to reach a desired lower speed for driving in a curve. Therefore, the measured TOrt is higher than the actual TOrt, as the driver was already using the steering wheel for lane keeping. Therefore, it is recommended to use a different calculation method for the TOrt based on the steering angle. For future experiments, the use of both locations is encouraged for distinguishing between low and high task demand, as evident workload differences were measured.

Task during automated driving

Based on the study by Merat et al. (2014), it was hypothesised that the monitoring task would result in less TOR-induced workload compared to driver workload when Tetris was played during automated driving. However, the RTLX measurements do only indicate a small workload difference between the two tasks, with 30.62% when monitoring and 31.99% when Tetris was played. This was contrary to expectations, as it was hypothesised that the driver would not except the TOR when being engaged in the Non-Driving Related Task (NDRT) during automate driving. However, this small workload difference can be attributed to the experimental setup with the researcher as the only participant, which had the disadvantage of reducing the 'surprise' effect of the TOR. For future studies that involve repeated participation in an experiment, it is recommended to have a variable duration of automated driving, to have a surprise effect of the TOR.

Mental demand as measured by the RTLX is the only scale that reports a higher value for the monitoring task. This suggests that playing Tetris during automated driving had a relaxing effect, easing mental demand felt during the experiment run.

Physiological workload as measured by HR also shows that playing Tetris during automated driving increased TOR-induced workload. HR increased on average by +1.52 HR after the TOR in scenarios with Tetris as NDRT, whereas an increase of +0.52 HR is reported in scenarios with a monitoring task. On the contrary, the RMSSD reports a decreased workload after the TOR for both tasks. Interesting is the reported decrease in SDNN after the TOR when monitoring (indicating increased workload), and an increase in SDNN when Tetris was played (indicating decreased workload). The expectation was that the monitoring task would reduce the effect of the take-over on workload. However, it can be hypothesised that monitoring could have resulted in drowsiness and inattention, which therefore increased the surprise effect of the TOR (Endsley, 2019). Also, because Tetris is a cognitive task the driver could be more aware of the time that had passed since the hand-over, which decreases the surprise effect of the TOR. The HR graph in Figure 5.2d supports the hypothesis, by showing a steep peak immediately after the TOR in scenarios with the monitoring task. Whereas in scenarios with Tetris as NDRT, a delayed and less steep increase in HR is observed after the TOR.

The TOrt in scenarios with Tetris as NDRT is longer than in scenarios with a monitoring task. This contradicts the hypothesis that monitoring resulted in drowsiness and inattention and therefore a slower reaction time (Endsley, 2019). However, the Tetris task necessitates the driver to reposition before being able to take-over, which increased the TOrt. Therefore it was expected that an exceedance of the time budget occurred more frequently in scenarios with Tetris as NDRT. However, this was not the case, even though the average exceedance is larger compared to scenarios with the monitoring task. Speed deviation and SDLP are similar for both tasks. The mean speed is also almost equal for the tasks. However, the speed at take-over is on average lower in scenarios with Tetris as NDRT. This is caused by a delayed acceleration, which is reflected by the larger average exceedance of the time budget in scenarios with Tetris as NDRT.

In summary, it was expected that a clear difference in the effect of the TOR on workload could be distinguished between the two tasks, i.e. when engaged in Tetris, the TOR was expected to have a greater effect on workload compared to the monitoring task. However, the results suggest that monitoring led to underload, resulting in a greater effect of the TOR on workload. Moreover, the results also suggest that playing Tetris did not resulted in underload, as was expected based on (Endsley, 2019). In the current experimental set-up, no clear differences in the effect of workload was measured between the two tasks. Possibly, if the duration of automated driving before the TOR was extended, playing Tetris would have resulted in underload, and therefore the TOR would have resulted in greater workload increases. However, due to the short duration of automated driving, playing Tetris did not result in drowsiness, rather it resulted in increased driver vigilance.

5.4 REFLECTION ON THE USED DRIVER WORKLOAD MEASURES

Subjective workload

Unique to this study is the trend analysis that provides insight into the development of workload with repeated take-overs. It was expected that increasing underload would develop during the experiment, which would be reflected in significant or near-significant workload trends. Namely, as the experiment required the participant to drive an almost identical driving scenario repeatedly in a short time span, fatigue and drowsiness were expected to develop. This would result in underload, that required the participant to increase effort in maintaining an equal performance in the experiment (Endsley and Kiris, 1995). A possible factor contributing to the non-significant trends can be attributed to respondent fatigue that occurred because of the high frequency of filling in the Raw Task Load Index (RTLX). This could have led to response bias, that includes memory bias and decreased consideration to the RTLX (Lavrakas, 2008). As a result, the RTLX would not reflect the true experienced workload.

Physiological workload

The negative slope of the mean Heart Rate (HR) (-0.10 beats per minute (bpm)), and positives slopes of the mean Root Mean Square of Successive Differences (RMSSD) (+0.15 ms) and Standard Deviation of Normal to Normal intervals (SDNN) (+0.12 ms) indicate decreasing workload as the experiment progresses. However, only the HR trend was found significant. Increasing HR and decreasing RMSSD and SDNN was reported in both consecutive runs (in the various parts) as all runs combined, this suggests the occurrence of a time-on-task effect. This same effect was also found in Heikoop et al. (2019a). Namely, HR and Heart

Rate Variability (HRV) are sensitive to variations in vigilance. With decreases in vigilance resulting in decreases in HR and increases in HRV.

A diminishing effect of the take-over on HR (-0.02 bpm) was found. Although non-significant this suggests that repeated take-overs decrease take-over request (TOR)-induced workload. TOR anticipation improves when the driver repeatedly experiences similar TORs. This can be attributed to the TOR, which is always issued at the same two locations after the same two durations of automated driving.

This study included a novel approach using the RMSSD and SDNN to measure TOR-induced workload. Currently, there is no consensus on the usefulness of these HRV measures as workload measure (Mehler et al., 2011; Luque-Casado et al., 2016; Hidalgo-Muñoz et al., 2019). As far as is known, only Pakdamanian et al. (2020) (published at the time of writing this study) used RMSSD and SDNN to measure TOR-induced workload. Pakdamanian et al. (2020) conducted an exploratory study with two participants who experienced four TORs under two weather conditions (sunny / rain) and alert modalities (visual-auditory / auditory). However, unfortunately their data cannot be translated to this study, as Pakdamanian et al. (2020) only presented the mean RMSSD and SDNN after the TOR instead of the respective increases or decreases after the TOR. Other studies used the RMSSD and SDNN to measure workload differences between low and high task demands during automated driving (Mehler et al., 2011; Luque-Casado et al., 2016; Hidalgo-Muñoz et al., 2019; Heine et al., 2017; Shakouri et al., 2018; Heikoop et al., 2018). Based on the results of this study, the SDNN appears to be a more sensitive workload measure than RMSSD, as RMSSD often indicated a different direction of the expected effect on workload for the various design variables. Future studies using the RMSSD and SDNN to measure specifically TOR-induced workload are necessary to gain a better understanding of the sensitivity of RMSSD and SDNN as (TOR-induced) workload measure, compared to other proven physiological workload measures such as HR.

Driving performance

It was expected that learning effects in driving performance would be found due to the repetitive nature of the experiment. However, no significant trends were found, nor do the found trends suggest the emergence of any learning effect. Previous studies that required participants to respond to multiple TORs did find learning effects, for instance Körber et al. (2016). As discussed earlier in Section 5.2, the take-over reaction time (TOrt) in 5-, 10- and 15-second time budget scenarios is longer than found in comparable studies due to the modality of resuming control before expiration of the time budget. Thus, finding no learning effect in TOrt entails that the duration between the TOR and pressing the key-combination on the keyboard remained constant.

Secondary task

An advantage of using Tetris as Non-Driving Related Task (NDRT) is that is assured engagement in the NDRT. However, it was hypothesised that Tetris scores would provide indication of NDRT engagement and, therefore, Tetris would be an appropriate workload measure. However, this hypothesis is rejected for multiple reasons. Firstly, a significant trend was found indicating that the participant was getting better at playing Tetris. Secondly, the significant difference in Tetris scores between the two TOR locations. Finding this difference was evident as the duration of automated driving (i.e. playing Tetris) differed between the two locations. When a TOR was issued in a curve, the duration of automated driving (playing Tetris) was shorter, thus lower Tetris scores could be expected. However, when designing the experiment and selecting appropriate workload measures,this downside of using Tetris as workload (task engagement) measure was overlooked. Thirdly, with a perfect design of the simulation, the Tetris score obtained during automated driving before the TOR should not give any indication on TOR-induced workload as the game is stopped once the TOR is issued. However, as discussed earlier in Section 3.2.9, one characteristic affecting TOR-induced workload was already revealed to the participant at the start of the simulation. This was the traffic density at the take-over, which could have affected NDRT engagement.

However, the results showed that there was no such relationship between the Tetris scores and traffic density (Section 5.3). The question is whether revealing this characteristic had no influence on the Tetris scores or if the Tetris scores were not sensitive enough to be able to measure distraction. The latter is expected, therefore for future studies it is recommended to not use Tetris as workload measure, but rather use the validated *n*-back task.

5.5 RECOMMENDATIONS FOR FUTURE EXPERIMENTS

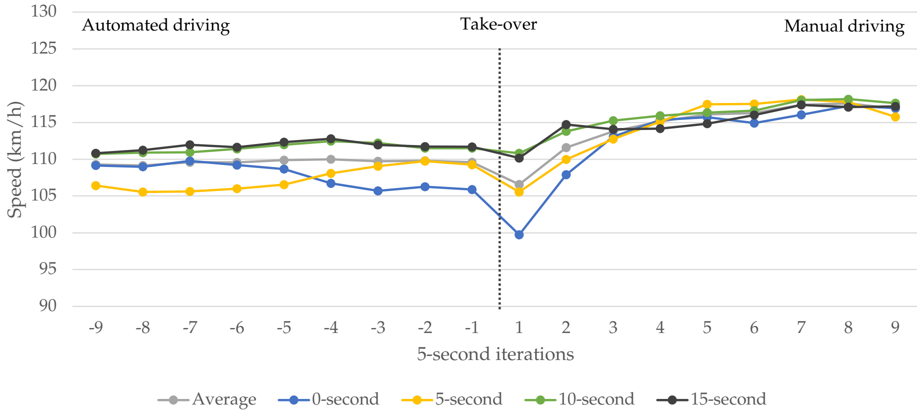
5.5.1 Design of the simulation

Some recommendations can be made for improving the current experimental design, which in its current state affected take-over request (TOR)-induced workload. First of all, it is recommended to design a variable duration of automated driving in future studies involving repeated TORs. In the current study the duration of automated driving was fixed for the two TOR locations. Therefore, the participant was able to anticipate the TOR, which mitigated the effect of the TOR on driver workload. This has not only affected the effect of Non-Driving Related Task (NDRT) engagement on TOR-induced workload, but has also increased the time-on-task effect which was measured in physiological workload.

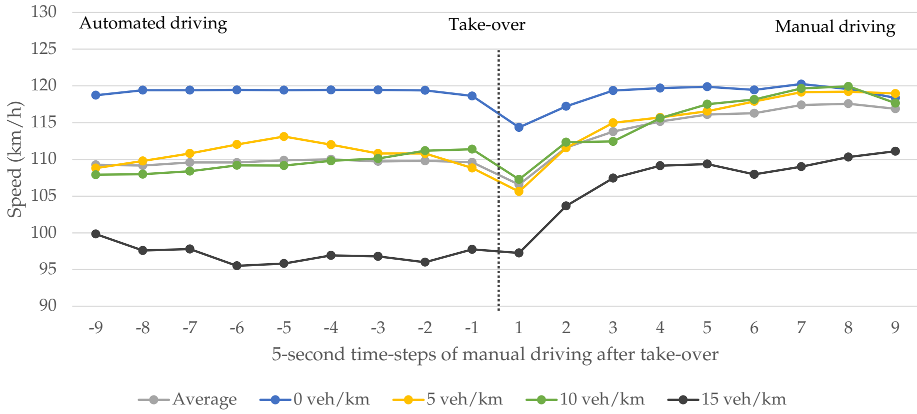
Second, it is recommended to reveal the traffic density at the take-over location. This has two reasons. Firstly, revealing the traffic density at the start of the experiment decreased the need to analyse the driving environment (which eases the demand of the TOR). Secondly, it resulted in a decreased speed at the take-over. As illustrated in Figure 5.1, initial speed at the take-over decreased with increases in traffic density (and cannot be attributed to other design variables). As discussed in Section 5.3, the various workload measures did not measure the greatest increase in workload in scenarios with a high traffic density of 15 vehicles/km/lane. This could be attributed to the lower initial speed in high traffic density scenarios. The limitation of revealing the traffic density at the start was already expected (Section 3.2.9), however, I was unable to design it differently in Unity (even with help of others). This will be a challenge for future researchers who will continue working on the simulation.

Third, this study found an increased take-over reaction time (TOrt), which can be attributed to the take-over modality. A key-combination had to be pressed on the keyboard, which resulted in an increased TOrt for scenarios with 5-, 10-, and 15-second time budgets. Again, before the experiment it was already known that the take-over modality would probably increase the TOrt. However, I was unable to design it differently in Unity. This will also be a challenge for future researchers who will continue working on the simulation.

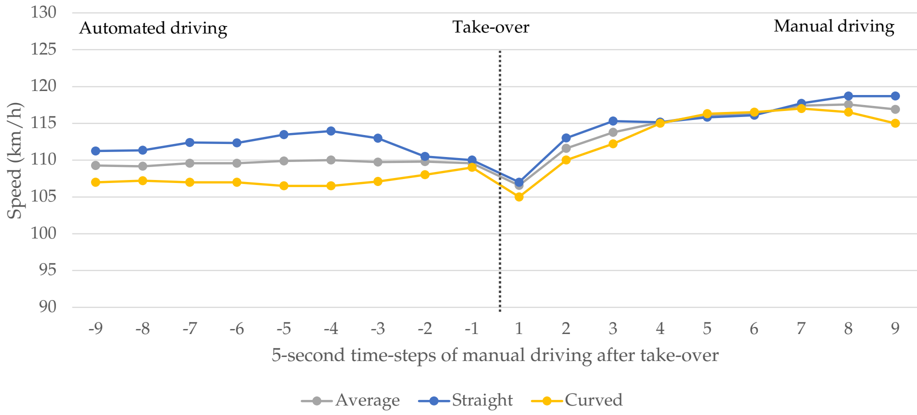
Fourth, it is recommended to increase the reality of the simulation by allowing flexible speeds and overtaking behaviour of other vehicles. At this point, I was not possible to allow this behaviour as it significantly deteriorated smooth-running of the simulation. However, by reducing the number of agents (vehicles), the speed of the simulation can be improved. This can be achieved by spawning vehicles directly behind the ego-vehicle and de-spawning them if they are more than e.g. 500 meters away from the ego-vehicle (when out of sight). However, I was unable to design this using Unity.



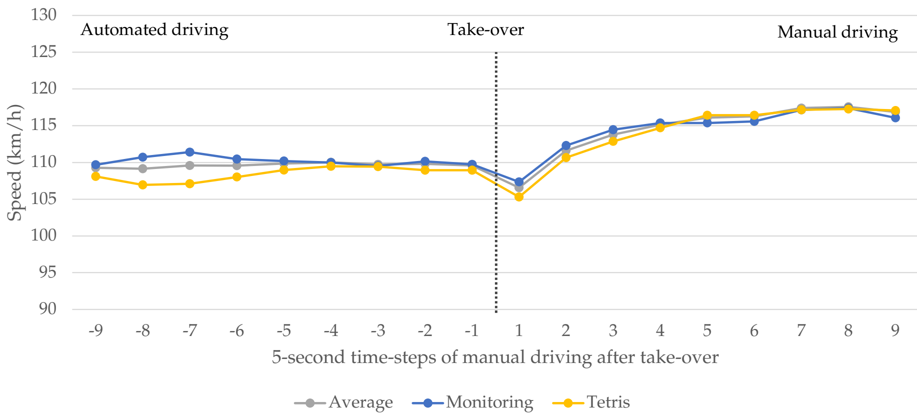
(a) Mean speed during manual driving after take-over for the different time budgets



(b) Mean speed during manual driving after take-over for the different traffic densities



(c) Mean speed during manual driving after take-over for the different TOR locations



(d) Mean speed during manual driving after take-over for the different NDRTs

Figure 5.1: Mean speed during manual driving after resumption of control (take-over) for the different conditions; 5-second time-steps are used to analyse 1 minute of manual driving

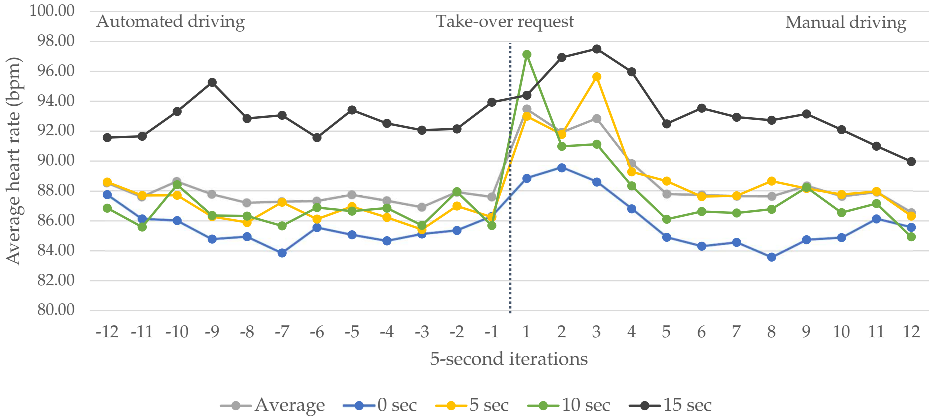
5.5.2 Apparatus

The finger sensor used for the physiological workload measurements regularly recorded ambient noise.The noise could be filtered out in most measurements. However, too much ambient noise was recorded in 14 scenarios, and therefore limited physiological workload data was available for these scenarios (i.e. only the mean Heart Rate (HR) was analysed making use of the chest strap data). Moreover, the measurements were therefore not of sufficient quality to measure to include High-Frequency (HF) power as Heart Rate Variability (HRV) workload measure. The finger sensor recorded too little R-to-R Intervals (RRIs) to reliably calculate frequency measurements. As a result, the calculated HF-power was far outside the expected value (Shaffer and Ginsberg, 2017; Gent, 2019). Moreover, usage of a finger sensor to measure physiological workload, as done in this study, is not a perfect match with a $N = 1$ study. Checking for correct placement of the finger sensor is only done before the run and is not monitored during the run, which could have prevented data collection issues. It is therefore recommended to make use of a notification when a certain level of ambient noise is being recorded or to make use of a different attachment method of which decreases noise recording. Using Electrocardiogram (ECG) equipment is also recommended, although more time-consuming than using a Photoplethysmography (PPG) finger sensor.

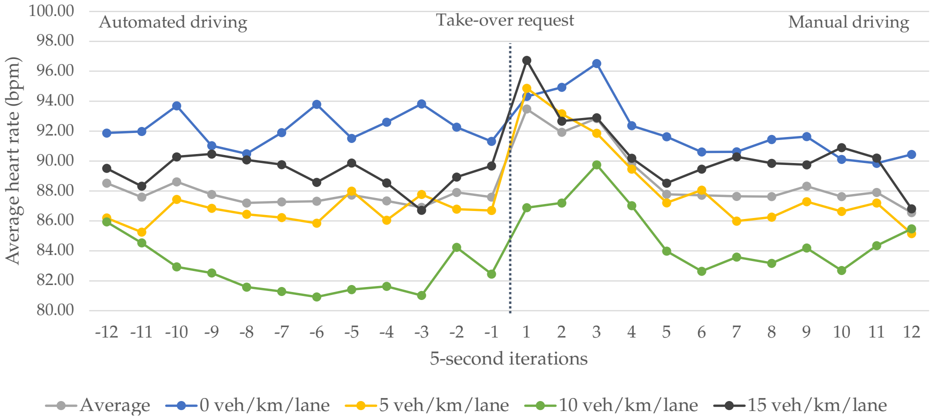
The following recommendation concerns the steering wheel of the driving simulation, which was unable to steer with the vehicle during automated driving. In other words, at the hand-over the steering wheel had a certain angle and remained at the same angle during automated driving. If at the hand-over the steering wheel was tilted to the right, and the TOR was issued in a curve, the driver was required to perform an excessive steering manoeuvre to remain in the current lane. This has possibly contributed to the evident and sometimes significant differences in TOR-induced workload between the two take-over locations. Therefore, it is recommended to use a driving simulator with a steering wheel that is able to follow the steering movements of the simulation during automated driving.

5.5.3 Driver workload analysis

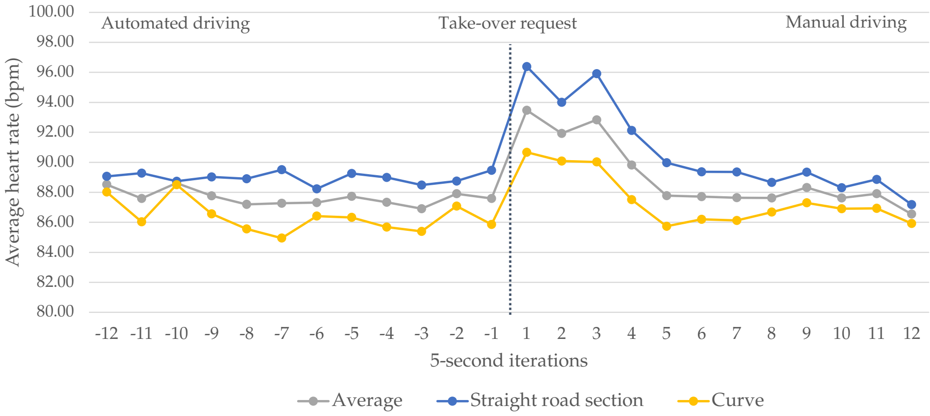
As elaborated in Section 3.2.2, the effect of the TOR on driver workload measured by the HR was determined over a one minute duration before and after the TOR. This has resulted in small effect sizes measured using the HR. However, it appeared that a one minute duration was too long for measuring the effect of the TOR on the driver workload. This is illustrated in Figure 5.2, which shows how HR is elevated up to 25 seconds after the TOR. If a 25-second duration was used, a larger effect size of the TOR on driver workload could be demonstrated.



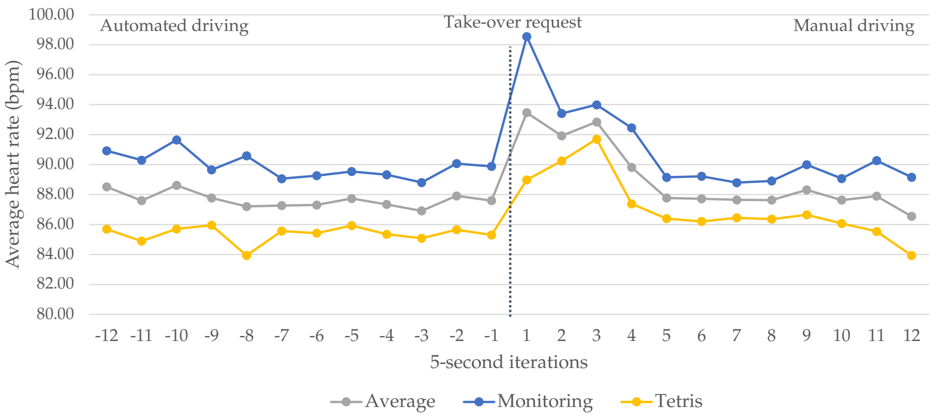
(a) Time budget



(b) Traffic density



(c) Location TOR



(d) Task during automated driving

Figure 5.2: HR course 1 minute before and after the TOR in different task load conditions

6 CONCLUSION

This research concerned validating the design of a driving simulator experiment regarding the effect of take-overs on workload and the role of personality. From the aim to study the role of personality in automated driving followed the research question of the study: *"To what extent does personality interact with driver workload induced by a take-over?"* A driving simulator experiment was designed to study the effect of the take-over on the workload that people with different personalities experience. In this study, the design of the driving simulator experiment was validated, and recommendations to improve the experiment were formulated. Three sub questions were formulated in order to answer the research question, which are answered in Section 6.1, 6.2, and 6.3. A concluding remark on the research is formulated in Section 6.4.

6.1 DESIGN OF THE EXPERIMENT

The first sub-question was: *"What is a suitable design of a driving simulation experiment in which the effect of a take-over on workload can be measured?"* The effect of the design of the simulation on workload was investigated to answer this question. For this reason, four design characteristics were varied that, based on literature, were expected to have a significant effect on workload. Variations of these design characteristics were tested in the simulation to determine which variation is best suited to use in the experiment. This experiment examined the effect on workload of different time budgets of the take-over request (TOR), traffic densities, the location of the TOR, and involvement in a Non-Driving Related Task (NDRT) during automated driving. Based on the results of the $N = 1$ experiment, recommendations were formulated on the use of the design variables in future experiments. Table 6.1 provides an overview of the design variables and attribute levels varied in the experiment and the respective recommendations on designs of future experiments.

Table 6.1: Recommendations on design variables in future experiments

Design variables	**Levels**	**Recommendations**
Time budget [s]	0	Low demand: 10- or 15-second time budget
	5	High demand: 0-second time budget
	10	
	15	
Traffic density [veh/km/lane]	0	Low demand: 0 vehicles/km/lane
	5	High demand: 5, 10 or 15 vehicles/km/lane
	10	
	15	
Location of the TOR	Straight road	Low demand: straight road
	Curve	High demand: curve
Task during automated driving	Monitoring	Short duration: monitoring
	Tetris	Long duration: monitoring or constant demand NDRT

Time budget

The various workload measures did not unambiguously indicate differences in TOR-induced workload with different time budgets. It was expected that every increase in time budget would result in a smaller effect of the TOR on workload . The Raw Task Load Index (RTLX) suggested, indeed, that workload decreases with a longer available duration to respond to the TOR. Similarly, the Standard Deviation of Lateral Position (SDLP) suggested that the 0-second time budget had a greater effect on workload than time budgets of 5 seconds or greater. However, driving performance as indicated by the SDLP did not seem to improve with a time budget duration greater than 5 seconds. The Standard Deviation of Normal to Normal intervals (SDNN) also showed how immediate disengagement of the Automated Driving System (ADS) resulted in greater workload compared to scenarios where a time budget was provided to the driver to respond to the TOR. So, to simulate an urgent (high demand) take-over, it is recommended to use a 0-second time budget instead of a 5-second time budget. Moreover, time budgets of 10 or 15 seconds are recommended for non-urgent take-overs, as based on the results of this study it is expected that 10- and 15-second time budgets will be significantly different from 0-second time budgets in future experiments involving participants.

Traffic density

It has been found that driving performance after the TOR decreased with increases in traffic density in the current experimental setup. This provides an indication of the emergence of overload when traffic density at the take-over increases. However, no overload has occurred in scenarios with a high traffic density, as indicated by subjective and physiological workload. With the current experiment setup, is is recommended to simulate either a low or medium traffic density of 5 or 10 vehicles per kilometre per lane for simulating a take-over at high demand. Namely, both low and medium traffic density resulted in a similar higher workload compared to zero traffic density. However, preference is given to simulate a medium traffic density, as already in this experiment setup significant workload differences were found between no-traffic and medium traffic density scenarios. For low demand take-overs, it is recommended to use zero traffic density instead of a low traffic density (similar workload low and medium traffic density). Simulating a high traffic density of 15 vehicles per kilometre per lane was not ideal in the current design of the simulation. The ego-vehicle remained on the right lane during automated driving. At the take-over, a lower initial speed is found compared to the other traffic densities, which could have resulted in less TOR-induced workload in high traffic density scenarios. High traffic density could be used for high demand take-over in future experiments if a lower speed during automated driving can be prevented in high traffic density scenarios.

Location of the take-over request (TOR)

In the current experiment setup, the TOR in a curve induced more workload compared to a TOR on a straight road section. The difference in workload was especially clear in the difference in subjective workload measured by the RTLX. Physiologically, the difference in workload was also measured by the Root Mean Square of Successive Differences (RMSSD) and SDNN. Post-take-over driving performance is noticeably worse when the TOR was issued in a curve, as measured by the speed deviation and SDLP. However, a distorted take-over reaction time (TOrt) is possibly found; the mean TOrt in scenarios with a TOR in a curve is possibly shorter than measured. This can be attributed to the method of TOrt calculation, which measures the time between the TOR and the first acceleration. As the speed of the vehicle at the take-over in a curve is approximately 120 km/h, the driver delays acceleration to reach a desired lower speed for driving in a curve. This therefore increases the calculated TOrt, although the driver was already using the steering wheel to keep driving on the current lane. Therefore, it is recommended to use a different calculation method for the TOrt based on the steering angle. For future experiments, the use of both locations is encouraged for distinguishing between low (straight road) and high (curve) task demand, as for both clear workload differences were measured.

Task during automated driving

It was expected that a clear difference in the effect of the TOR on workload could be distinguished between the two tasks, i.e. when engaged in Tetris, the TOR was expected to have a greater effect on workload compared to the monitoring task. However, in the current experiment setup not a clear difference in the effect of workload was measured between the two tasks. RTLX measurements were similar as well as driving performance, except for the TOrt which was higher when Tetris was played. Physiological workload does not provide a clear indication: Heart Rate (HR) suggests that the TOR induced more workload when Tetris was played, RMSSD indicates a decrease in workload after the TOR and more so when monitoring, and SDNN indicates that workload decreased after the TOR when Tetris was played but increased greatly when monitoring. SDNN as Heart Rate Variability (HRV) workload measures appeared to be more sensitive to variations in task demand at the TOR, based on the workload findings for the different time budgets, traffic densities and TOR locations. Therefore, with an assumed reliable SDNN, a clear physiological workload difference distinguishes the two tasks. It is therefore expected that monitoring resulted in out-of-the-loop issues, which increased TOR-induced workload. These out-of-the-loop issued were also expected to emerge when playing Tetris. However, the short duration of playing Tetris during automated driving resulted in sustained vigilance, whereas monitoring decreased vigilance. Possibly, if the duration of automated driving before the TOR was extended, vigilance would decrease when playing Tetris. Therefore, in order to simulate out-of-the-loop driver' issues, it is recommended to use monitoring as task during automated driving when the duration of automated driving is short (i.e. up to 3:00 minutes). With a longer duration of automated driving, it is expected that vigilance would also decrease when playing Tetris. However, Tetris is not an appropriate NDRT for a long duration of automated driving as task demand increases during the game. Instead, only a NDRT with a constant task demand is an appropriate NDRT for a longer duration of automated driving (i.e. an educational movie or book).

Reflection on the design

Recommendations were made to include specific designs of the driving simulation experiment for simulating low and high demand take-overs. If either a low or high demand is simulated in a future experiment, this will not provide a complete picture of automated driving behaviour for different personalities. Namely, a take-over at low or high demand can result in different behaviours for different personalities. For example, high demand, on the one hand, could result in overload quicker for some personalities, which is detrimental for driving safety at take-overs. Low demand, on the other hand, could cause inattention to the driving environment for people with other personality traits, which decreases driving safety at take-over. People differing in personality traits will therefore probably exhibit different behaviour at different task demands. For this reason, it is recommended to simulate both low and high task demand at the take-over in a future driving simulation experiment. It can be decided to have participants drive two simulation runs differing in task demand in a counterbalanced order.

By combining the recommended variables to design low or high demand at the TOR, variations in task demand can be obtained. However, for a realistic take-over, the urgency of the time budget must be in accordance to the critically of the take-over situation (Eriksson and Stanton, 2017). For this reason, it is recommended to use the following design variables for a TOR with low task demand (in the current experiment setup): issuing a TOR on a straight stretch of the road with no traffic with a provided time budget of 10/15 seconds. On the contrary, for high task demand, a TOR can be issued when driving on a curve with medium/high traffic density with immediate disengagement of the simulated ADS (0-second time budget).

6.2 interaction between personality and workload

The second sub-question formulated for this study was: *"How can the interaction between personality and workload be investigated in a driving simulation experiment regarding the effect of a take-over on workload?"* Besides design recommendations for studying the interaction between personality and workload, recommendations are also formulated on the method of measuring workload. In this study, workload was measured in four ways, namely subjectively, physiologically, based on driving performance and secondary task performance. Subjective workload was measured using the Raw Task Load Index (RTLX), which is also recommended for future studies. For measuring physiological workload, both a Photoplethysmography (PPG) finger sensor and chest strap were used. The finger sensor was sensitive to measuring ambient noise, which was filtered out in most measurements to accurately measure Heart Rate (HR), Root Mean Square of Successive Differences (RMSSD), and Standard Deviation of Normal to Normal intervals (SDNN). However, too much ambient noise was recorded in 14 scenarios, and therefore limited physiological workload data was available for these scenarios (i.e. only the mean HR was analysed making use of the chest strap data). It was intended to include High-Frequency (HF) power as time-frequency Heart Rate Variability (HRV) measure, however, although noise could filtered in must runs, the data was not of sufficient quality to accurately measure HF power. Namely, the finger sensor is sensitive to record noise especially at the take-over as the hand is moved from a resting position or the tablet to the steering wheel. The finger sensor could be an ideal method for measuring workload, if movement of the sensor can be prevented, even at a take-over. Especially as using the finger sensor over an Electrocardiogram (ECG) measurement tool is preferred under the COVID-19 circumstances, as it minimises contact between the researcher and the participant.

This study included a novel approach using the RMSSD and SDNN to measure take-over request (TOR)-induced workload. There is no consensus yet on the usefulness of these HRV measures as workload measure. As far as is known, only Pakdamanian et al. (2020) (published at the time of writing this study) used RMSSD and SDNN to measure TOR-induced workload. Pakdamanian et al. (2020) conducted an exploratory study with two participants who experienced four TORs under two weather conditions and alert modalities. Other studies used the RMSSD and/or SDNN to measure workload differences between low and high task demands during automated driving, although not in relation to TORs. Based on the results of this study, the SDNN appears to be a more sensitive workload measure than RMSSD, as RMSSD often indicated a different direction of the expected effect on workload for the various design variables. As the results of this study are not conclusive about the usefulness of these measures, it is advised for the future study regarding personality and other future studies to include both measures in order to obtain a better understanding of the validity of RMSSD and SDNN as workload measure in general and in relation to take-overs.

Regarding driving performance measures, it is advised to use the take-over reaction time (TOrt), speed deviation and Standard Deviation of Lateral Position (SDLP) in a future experiment with the same apparatus. Other robust driving performance measures, such as the maximum steering wheel angle and the number of 1° steering wheel reversals, were not used as workload measure in the experiment, because the steering wheel of the driving simulation was very sensitive to small movements in the wheel. Small movements on the steering wheel sometimes led to excessive unintended steering manoeuvres. Therefore, the TOrt is defined as the duration between the TOR and the first acceleration. Small unintentional movements of the steering wheel could have appeared as a driver take-over manoeuvre. To take-over manual driving before expiration of the time budget it was required to press a key combination on the keyboard which was placed on the dashboard of the driving simulator. As pressing this key combination could have increased the TOrt, it is recommended for future experiments to use a button on the steering wheel instead, or to allow overruling of the system when the steer, gas pedal, or brake is used.

Tetris was used as both a Non-Driving Related Task (NDRT) and as secondary task performance measure. However, it was found that the Tetris scores did not reflect driver workload during the experiment, instead it was a perfect example of a learning curve that occurred in the experiment. Therefore making it unsuitable as a secondary task performance measure. For this study it was chosen to include several workload measures in order to gain a complete understanding of workload under different task demands. Tetris was chosen over other tasks because it was suitable to use in a $N = 1$ experiment and by its low complex data analysis. However, for future studies involving participants (or other $N = 1$ experiments) it is recommended to use a n-back task instead, which is a proven workload measure.

6.3 EFFECT OF TAKE-OVERS ON WORKLOAD IN FUTURE EXPERIMENT

The third and final sub-question is: *"What effect of take-overs on workload can be expected in a study involving participants, based on the results of this self-experiment?"* Based on this self-experiment, no hypotheses can be made about how workload is experienced by people with different personalities. Because of researcher bias in combination with my personality, which is not predominantly dominant in one trait (dominant personality was agreeableness with 35pt compared to the mean of 41.55pt, Table 4.1), little can be hypothesised on the agreeableness trait in automated driving.

Based on the self-experiment, it can be expected that learning effects and time-on-task effects will occur when participants experience multiple simulation runs. In the present study, time-on-task effects were found in the physiological workload because of decreased vigilance as the experiment progressed. The results of the study also showed that take-over request (TOR)-induced workload decreased as the experiment progressed. A decreasing effect of the TOR on workload in consecutive runs provides indication of a time-on-task-effect, whereas the decreasing effect of the TOR on workload between experiment days indicates an acclimatisation effect to the experimental setting. Moreover, the results of the current study suggest that a learning effect occurred in the take-over reaction time (TOrt). Learning effects were also expected for speed deviation and Standard Deviation of Lateral Position (SDLP), lack thereof can be attributed to research bias. Prerequisite knowledge about the simulation could have improved driving performance. The researcher knows the experiment by heart, i.e. the location of the TOR and the required action at the take-over. The learning curve of the researcher started before the experiment, complicating finding significant improvements in driving performance. If a participant without prerequisite knowledge was the test subject, a learning curve in driving performance could have been found.

Based on the $N = 1$ experiment, reference workload values are provided for the aforementioned TORs at low and high demand. These reference values can be valuable for the future experiment regarding personality in automated driving. However, it is important to note that these values should only be used as guideline and not as golden standard for measuring workload in the proposed experimental setting. The reference values are tabulated in Table 6.2. The average is based on the 63 scenarios which were run in the experiment (one was accidentally not tested). The reference values for low and high demand are based on the average workload which was measured for every separate attribute level pertaining to low or high demand. For low demand this entails: the average of all scenarios with a time budget of 15 seconds, all scenarios without traffic, and all scenarios with a TOR issued on a straight stretch of the road. For high demand: the average of all scenarios with a time budget of 0 seconds, all scenarios with medium traffic, and all scenarios with a TOR issued when driving on a curve.

Table 6.2: Reference values TOR-induced workload, based on $N = 1$ experiment

Workload measure	Average	Low demand	High demand
Subjective workload			
RTLX overall workload [%]	31.29	22.97	40.10
Physiological workload			
HR [bpm]	+ 1.02	+ 0.54	+ 1.90
RMSSD [ms]	+ 0.31	- 0.96	- 0.90
SDNN [ms]	- 2.26	- 1.67	- 4.67
Driving performance			
TOrt [s]	A different calculation method and take-over modality is proposed		
SD speed [km/h]	5.19	4.52	5.62
SDLP [cm]	25.59	21.42	27.26
Secondary task performance			
Tetris	Not recommended as workload measure		

6.4 THE ROLE OF PERSONALITY

Unfortunately, due to the COVID-19 circumstances the main research question (*"To what extent does personality interact with driver workload induced by a take-over?"*) remains open for the future study with over 100 participants regarding personality in automated driving. By answering the different sub-questions, this research provided recommendations for the personality research, not only limited to the design of the driving simulation research, but also how the experiment can be conducted and analysed.

All in all, conducting an $N = 1$ study proved valuable for validating the design of the driving simulation experiment. In light of the COVID-19 conditions that complicated conducting studies involving participants, it proved to be a valuable research method. Besides recommendations on improvement of the design of the driving simulation experiment, the results of the experiment provided indication of usefulness of including the Standard Deviation of Normal to Normal intervals (SDNN) as physiological workload measure. In the absence of the need of including participants other than the researcher, the $N = 1$ study proved to be a more accessible and easily applicable method of validating research set-ups. Based on the results of a $N = 1$ study, indications can be formulated as to whether or not certain research results can be formulated which allows a more focused approach in future studies.

BIBLIOGRAPHY

Adminaite-Fodor, D., Heilpern, C., and Jost, G. (2019). Ranking EU progress on road safety: 13th road safety performance index report. Technical report, European Transport Safety Council.

Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3):1–39.

Alrefaie, M. T., Summerskill, S., and Jackon, T. W. (2019). In a heart beat: Using driver's physiological changes to determine the quality of a takeover in highly automated vehicles. *Accident Analysis and Prevention*, 131:180–190.

Avram, R., Tison, G. H., Aschbacher, K., Kuhar, P., Vittinghoff, E., Butzner, M., Runge, R., Wu, N., Pletcher, M. J., Marcus, G. M., and Olgin, J. (2019). Real-world heart rate norms in the Health eHeart study. *NPJ Digital Medicine*, 2(58):1–10.

Baek, S. J., Yang, J. H., and Yun, H. (2019). How do humans respond when automated vehicles request an immediate vehicle control take-over? In *Adjunct Proceedings - 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019*, pages 341–345, New York, NY, USA. Association for Computing Machinery, Inc.

Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6):775–779.

BMW, FCA, Aptiv, Here, Audi, Baidu, Continental, Daimler, Intel, Volkswagen, and Infineon (2019). Safety first for automated driving.

Bueno, M., Dogan, E., Selem, F. H., Monacelli, E., Boverie, S., and Guillaume, A. (2016). How different mental workload levels affect the take-over control after automated driving. In *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, pages 2040–2045. IEEE.

Cattell, R. (1957). *Personality and motivation structure and measurement*. Harcourt, Brace, Jovanovich, New York.

Cho, C.-H., Cho, J., Baek, H. J., and Woo, J.-M. (2015). Reliability of ultra-short-term analysis as a surrogate of standard 5-Min analysis of heart rate variability. *Telemedicine and e-Health*, 21(5):404–414.

ChoiceMetrics (2018). Ngene user manual.

Clark, H. and Feng, J. (2017). Age differences in the takeover of vehicle control and engagement in non-driving-related activities in simulated driving with conditional automation. *Accident Analysis and Prevention*, 106:468–479.

Coughlin, J. F., Reimer, B., and Mehler, B. (2009). Driver wellness, safety & the development of an awarecar. Technical report, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Council of the European Union (2017). Valletta declaration on road safety. Technical report, Council of the European Union.

Dambock, D., Weissgerber, T., Kienle, M., and Bengler, K. (2013). Requirements for cooperative vehicle guidance. In *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, pages 1656–1661, The Hague. IEEE.

De Waard, D. (1996). *The measurement of drivers' mental workload*. Groningen University, Traffic Research Center Netherlands, Glimmen.

De Winter, J., Stanton, N. A., De Winter, J. C. F., Stanton, N. A., Price, J. S., Mistry, H., and Stanton, N. (2016). The effects of driving with different levels of unreliable automation on self-reported workload and secondary task performance. *International Journal of Vehicle Design*, 70(4):297:324.

De Winter, J. C., Happee, R., Martens, M. H., and Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour*, 27(PB):196–217.

De Winter, J. C. F., Van Leeuwen, P. M., and Happee, R. (2012). Advantages and disadvantages of driving simulators: A discussion. In Spink, A., Grieco, F., Krips, O., Loijens, L., Noldus, L., and Zimmerman, P., editors, *8th International Conference on Methods and Techniques in Behavioral Research*, pages 47–50, Utrecht, The Netherlands. Noldus.

Denissen, J. J., Geenen, R., Van Aken, M. A., Gosling, S. D., and Potter, J. (2008). Development and validation of a Dutch translation of the Big Five Inventory (BFI). *Journal of Personality Assessment*, 90(2):152–157.

DeYoung, C. G., Peterson, J. B., and Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences*, 33(4):533–552.

Dukic Willstrand, T., Anund, A., Pereira Cocron, M., Griesche, S., Strand, N., Troberg, S., and Zanovello, L. (2018). Collecting end-users needs regarding driver state-based automation in the ADAS&ME project. In *Proceedings of 7th Transport Research Arena TRA 2018*, pages 1–10, Vienna, Austria.

Endsley, M. R. (2019). Situation awareness in future autonomous vehicles: Beware of the unexpected. In Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T., and Fujita, Y., editors, *Proceedings of the 20th Congress of the International Ergonomics Association*, volume 824, pages 303–309. Springer, Cham.

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

Eriksson, A. and Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors*, 59(4):689–705.

European Commission (2017). Europe on the Move: Commission takes action for clean, competitive and connected mobility — Mobility and Transport.

European Commission (2019). Guidelines on the exemption procedure for the EU approval of automated vehicles. Technical report, European Commission.

Eysenck, M. W., Derakshan, N., Santos, R., and Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, 7(2):336–353.

Feldhütter, A., Schneider, S., and Bengler, K. (2016). How the duration of automated driving influences take-over performance and gaze behavior. In *Arbeit in komplexen Systemen - Digital, vernetz, human?! 62. Kongress der Gesellschaft für Arbeitswissenschaft*, pages 309–318.

Fox, E. (2008). *Emotion science - Cognitive and neuroscientific approaches to understanding human emotions*. Palgrave Macmillan, New York.

Gent, P. V. (2019). Python heart rate analysis toolkit documentation.

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

Gold, C., Happee, R., and Bengler, K. (2018). Modeling take-over performance in level 3 conditionally automated vehicles. *Accident Analysis and Prevention*, 116:3–13.

Gold, C., Körber, M., Lechner, D., and Bengler, K. (2016). Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density. *Human factors*, 58(4):642–652.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of personality and social psychology*, 59(6):1216–1229.

Goldberg, L. R. (1992). Development of Factors for Big Five Factor Structure. *Psychological Assessment*, 4(1):26–42.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society*, pages 904–908.

Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52(C):139–183.

Heikoop, D. D., de Winter, J. C., van Arem, B., and Stanton, N. A. (2018). Effects of mental demands on situation awareness during platooning: A driving simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour*, 58:193–209.

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

Heikoop, D. D., Hagenzieker, M., Mecacci, G., Calvert, S., Sio, F. S. D., and Arem, B. v. (2019b). Human behaviour with automated driving systems: a quantitative framework for meaningful human control. *Theoretical Issues in Ergonomics Science*.

Heine, T., Lenis, G., Reichensperger, P., Beran, T., Doessel, O., and Deml, B. (2017). Electrocardiographic features for the measurement of drivers' mental workload. *Applied Ergonomics*, 61:31–43.

Hidalgo-Muñoz, A. R., Béquet, A. J., Astier-Juvenon, M., Pépin, G., Fort, A., Jallais, C., Tattegrain, H., and Gabaude, C. (2019). Respiration and heart rate modulation due to competing cognitive tasks while driving. *Frontiers in Human Neuroscience*, 12(525):1–8.

Hock, P., Kraus, J., Babel, F., Walch, M., Rukzio, E., and Baumann, M. (2018). How to design valid simulator studies for investigating user experience in automated driving - Review and hands-on considerations. In *Proceedings - 10th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2018*, pages 105–117. Association for Computing Machinery, Inc.

Inagaki, T. and Sheridan, T. B. (2019). A critique of the SAE conditional driving automation definition, and analyses of options for improvement. *Cognition, Technology and Work*, 21(4):569–578.

Jo, E., Lewis, K., Directo, D., Kim, M. J. Y., and Dolezal, B. A. (2016). Validation of biofeedback wearables for photoplethysmographic heart rate tracking. *Journal of Sports Science and Medicine*, 15(3):540–547.

John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In Pervin, L., editor, *Handbook of personality: Theory and research*, pages 66–100. Guilford Press, New York.

John, O. P., Naumann, L. P., and Soto, C. J. (2008). Paradigm shift to the integrative Big Five Trait taxonomy: History, measurement, and conceptual issues. In John, O., Robins, R., and Pervins, L., editors, *Handbook of personality: Theory and research*, pages 114–158. Guilford Press, New York, New York, USA.

John, O. P. and Srivastava, S. (1999). The Big Five trait taxonomy: history, measurement and theoretical perspectives. In *Handbook of personality: Theory and research*, chapter 4, pages 102–138. Guilford, 2nd edition.

Kerschbbaum, P., Lorenz, L., and Bengler, K. (2015). A transforming steering wheel for highly automated cars. In *2015 IEEE Intelligent Vehicles Symposium (IV) J*, pages 1287–1292.

Körber, M. and Bengler, K. (2014). Potential individual differences regarding automation effects in automated driving. In *INTERACCIÓN 2014: Proceedings of the XV International Conference on Human Computer Interaction*, pages 1–7, Puerto de la Cruz, Tenerife, Spain.

Körber, M., Gold, C., Lechner, D., and Bengler, K. (2016). The influence of age on the take-over of vehicle control in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 39:19–32.

Körber, M., Weißgerber, T., Kalb, L., Blaschke, C., and Farid, M. (2015). Prediction of take-over time in highly automated driving by two psychometric tests. *DYNA*, 82:195–201.

Lavrakas, P. (2008). Respondent fatigue. In *Encyclopedia of survey research methods2*. SAGE Publications.

Levin, S. and Wong, J. C. (2018). Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian — Technology — The Guardian.

Lohani, M., Payne, B. R., and Strayer, D. L. (2019). A review of psychophysiological measures to assess cognitive states in real-world driving. *Frontiers in Human Neuroscience*, 13(March):1–27.

Lorenz, L., Kerschbaum, P., and Schumann, J. (2014). Designing take over scenarios for automated driving: How does augmented reality support the driver to get back into the loop? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pages 1681–1685. SAGE Publications.

Lu, Z. and de Winter, J. C. (2015). A Review and Framework of Control Authority Transitions in Automated Driving. *Procedia Manufacturing*, 3:2510–2517.

Lu, Z., Happee, R., Cabrall, C. D., Kyriakidis, M., and De Winter, J. C. (2016). Human factors of transitions in automated driving: A general framework and literature survey. *Transportation Research Part F: Traffic Psychology and Behaviour*, 43:183–198.

Luque-Casado, A., Perales, J. C., Cárdenas, D., and Sanabria, D. (2016). Heart rate variability and cognitive processing: The autonomic response to task demands. *Biological Psychology*, 113:83–90.

Marberger, C., Mielenz, H., Naujoks, F., Radlmayr, J., Bengler, K. J., and Wandtner, B. (2017). Understanding and appplying the concept of "driver availability" in automated driving. In Stanton, N. A., editor, *Proceedings of the AHFE 2017 International Conference on Human Factors in Transportation*, pages 595–605. Springer.

Markus, H. R. and Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2):224–253.

McCrae, R. R. and Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52(1):81–90.

McCrae, R. R. and Costa, P. T. J. (1985). Updating Norman's "Adequate taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Article in Journal of Personality and Social Psychology*.

Mehler, B., Reimer, B., and Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: An on-road study across three age groups. *Human Factors*, 54(3):396–412.

Mehler, B., Reimer, B., Coughlin, J. F., and Dusek, J. A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record*, 2138(1):6–12.

Mehler, B., Reimer, B., and Wang, Y. (2011). A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. In *Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, pages 590–597, Olympic Valley, Lake Tahoe, California, Iowa City. IA Public Policy Center.

Melcher, V., Rauh, S., Diederichs, F., Widlroither, H., and Bauer, W. (2015). Take-over requests for automated driving. *Procedia Manufacturing*, 3:2867–2873.

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

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

Mok, B., Johns, M., Lee, K. J., Miller, D., Sirkin, D., Ive, P., and Ju, W. (2015). Emergency, Automation Off: Unstructured Transition Timing for Distracted Drivers of Automated Vehicles. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pages 2458–2464, Las Palmas. IEEE.

Munoz, M. L., Van Roon, A., Riese, H., Thio, C., Oostenbroek, E., Westrik, I., De Geus, E. J., Gansevoort, R., Lefrandt, J., Nolte, I. M., and Snieder, H. (2015). Validity of (ultra-)short recordings for heart rate variability measurements. *PLoS ONE*, 10(9):1–15.

Musek, J. (2007). A general factor of personality: Evidence for the Big One in the five-factor model. *Journal of Research in Personality*, 41(6):1213–1233.

NASA (2016). NASA-TLX.

National Highway Traffic Safety Administration (2008). National motor vehicle crash causation survey: Report to congress. Technical report, U.S. Department of Transportation.

Naujoks, F., Befelein, D., Wiedemann, K., and Neukum, A. (2017). A review of non-driving-related tasks used in studies on automated driving. In Stanton, N., editor, *Proceedings of the AHFE 2017 International Conference on Human Factors in Transportation*, pages 525–537. Springer, Cham.

Naujoks, F., Höfling, S., Purucker, C., and Zeeb, K. (2018). From partial and high automation to manual driving: Relationship between non-driving related tasks, drowsiness and take-over performance. *Accident Analysis and Prevention*, 121:28–42.

Naujoks, F., Mai, C., and Neukum, A. (2014). The effect of urgency of take-over requests during highly automated driving under distraction conditions. In Ahram, T., Karwowski, W., and Marek, T., editors, *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE*, volume 7, pages 431–438, Kruków, Poland.

Pakdamanian, E., Sheng, S., Coan, J. A., and Feng, L. (2020). Toward minimum startle after take-over request: A preliminary study of physiological data. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '20 Adjunct)*, pages 27–30, New York, NY, USA. ACM.

Paxion, J., Galy, E., and Berthelon, C. (2014). Mental workload and driving. *Frontiers in Psychology*, 5(1344):1–11.

Payre, W., Cestac, J., and Delhomme, P. (2016). Fully automated driving: Impact of trust and practice on manual control recovery. *Human Factors*, 58(2):229–241.

Peterkova, A. and Stremy, M. (2015). The raw ECG signal processing and the detection of QRS complex. In *IEEE European Modelling Symposium*, pages 80–85.

Petermeijer, S. M., De Winter, J. C., and Bengler, K. J. (2016). Vibrotactile displays: A survey with a view on highly automated driving. *IEEE Transactions on Intelligent Transportation Systems*, 17(4):897–907.

Polar (2020). Polar H10.

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., and Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. *Proceedings of the Human Factors and Ergonomics Society*, 2014-Janua(1):2063–2067.

Reimer, B., D'Ambrosio, L. A., and Coughlin, J. F. (2007). Secondary analysis of time of day on simulated driving performance. *Journal of Safety Research*, 38(5):563–570.

Reimer, B., Mehler, B., Wang, Y., and Coughlin, J. F. (2012). A field study on the impact of variations in short-term memory demands on drivers' visual attention and driving performance across three age groups. *Human Factors*, 54(3):454–468.

Riener, A., Jeon, M., Alvarez, I., and Frison, A. K. (2017). Driver in the loop: Best practices in automotive sensing and feedback mechanisms. In Meixner, G. and Müller, C., editors, *Automotive user interfaces*, pages 295–323. Springer, Cham.

Rijkswaterstaat (2019). Richtlijn ontwerp autosnelwegen 2019: ROA2019.

Rubio, S., Díaz, E., Martín, J., and Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology*, 53(1):61–86.

Ruscio, D., Bos, A. J., and Ciceri, M. R. (2017). Distraction or cognitive overload? Using modulations of the autonomic nervous system to discriminate the possible negative effects of advanced assistance system. *Accident Analysis and Prevention*, 103:105–111.

SAE International (2018). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Technical report, SAE International, Washington, DC.

Scharfe, M. S. L., Zeeb, K., and Russwinkel, N. (2020). The impact of situational complexity and familiarity on takeover quality in uncritical highly automated driving scenarios. *Information*, 11(115):1–12.

Schömig, N., Hargutt, V., Neukum, A., Petermann-Stock, I., and Othersen, I. (2015). The interaction between highly automated driving and the development of drowsiness. *Procedia Manufacturing*, 3:6652–6659.

Schöpplein, S. E. (2013). *Integration fahrstreifenbezogener Kenngrößen und seitlicher Detektionsdaten in ein makroskopisches Verkehrsflussmodell für dreistreifige Richtungsfahrbahnen*. PhD thesis, Technische Universität Munchen.

Selvaraj, N., Jaryal, A., Santhosh, J., Deepak, K. K., and Anand, S. (2008). Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. *Journal of Medical Engineering and Technology*, 32(6):479–484.

Shaffer, F. and Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5(258):1–17.

Shakouri, M., Ikuma, L. H., Aghazadeh, F., and Nahmens, I. (2018). Analysis of the sensitivity of heart rate variability and subjective workload measures in a driving simulator: The case of highway work zones. *International Journal of Industrial Ergonomics*, 66:136–145.

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

Task Force of the European Society of Cardiology and Electrophysiology (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17:354–381.

Taubman-Ben-Ari, O., Mikulincer, M., and Gillath, O. (2004). The multidimensional driving style inventory - scale construct and validation. *Accident Analysis and Prevention*, 36(3):323–332.

Taubman-Ben-Ari, O. and Yehiel, D. (2012). Driving styles and their associations with personality and motivation. *Accident Analysis and Prevention*, 45:416–422.

Tesla (2020). Model 3 owner's manual europe.

Triandis, H. C. (1989). The study of the self has a long tradition in psychology. *Psychological Review*, 96(3):506–520.

Van den Beukel, A. P. and Van der Voort, M. C. (2013). The influence of time-criticality on situation awareness when retrieving human control after automated driving. In *Proceedings of the 16th international IEEE annual conference on Intelligent Transport Systems*, pages 2000–2005, The Hague, The Netherlands.

Van Gent, P., Farah, H., Van Nes, N., and Van Arem, B. (2017). Towards real-time, nonintrusive estimation of driver workload: A simulator study. In *Proceedings of the Road Safety and Simulation Conference*, pages 1–10.

van Gent, P., Farah, H., van Nes, N., and van Arem, B. (2019). Analysing noisy driver physiology real-time using off-the-shelf sensors: Heart rate analysis software from the Taking the Fast Lane Project. *Journal of Open Research Software*, 7.

Varotto, S. F. (2018). *Driver behaviour during control transitions between adaptive cruise control and manual driving: Empirics and models*. TRAIL Research School.

Walch, M., Lange, K., Baumann, M., and Weber, M. (2015). Autonomous driving. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 11–18. Association for Computing Machinery (ACM).

Warm, J. S., Parasuraman, R., and Matthews, G. (2008). Vigilance requires hard mental work and is stressful.

Waymo LLC (2020). Journey.

Wright, T. J., Samuel, S., Borowsky, A., Zilberstein, S., and Fisher, D. L. (2016). Experienced drivers are quicker to achieve situation awareness than inexperienced drivers in situations of transfer of control within a Level 3 autonomous environment. In *Proceedings of the Human Factors and Ergonomics Society*, pages 270–273. Sage Publications.

Yoon, S. H. and Ji, Y. G. (2019). Non-driving-related tasks, workload, and takeover performance in highly automated driving contexts. *Transportation Research Part F*, 60:620–631.

Young, M. S. and Stanton, N. A. (2002). Attention and automation: New perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science*, 3(2):178–194.

Zeeb, K., Buchner, A., and Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accident Analysis and Prevention*, 78:212–221.

Zeeb, K., Buchner, A., and Schrauf, M. (2016). Is take-over time all that matters? The impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident Analysis and Prevention*, 92:230–239.

A | JOURNAL ARTICLE

Design of a driving simulation experiment regarding the effect of take-overs on driver workload: Implications of an N = 1 study

T. Marfoglia[a], D.D. Heikoop[a], J.C.F. de Winter[b], M.P. Hagenzieker[a]

[a]*Department of Transport & Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology*
[b]*Department of Cognitive Robotics, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology*

Abstract

The development of automated vehicles on the road is in full swing. As vehicles are getting increasingly automated, the human factor is diminished or eventually removed from automated driving. Until then, a combination of human input and automation is necessary during automated driving. This research focuses on the interaction between humans and machine and how a safe interaction can be designed by incorporating meaningful human control. Initially, the aim was to study how different personalities are reflected in driver workload induced by take-over requests (TORs). However, the COVID-19 circumstances changed the aim to validate the design of the driving simulation experiment by means of an $N = 1$ experiment. Design variables that have been found to play a role in driver workload are varied in the validation experiment. These variables are the duration of the time budget, traffic density, location of the TOR and task involvement during automated driving. Subsequently, workload was measured by a combination of subjective and physiological indicators and driving performance. Notably, this study includes the Root Mean Square of Successive Differences (RMSSD) and Standard Deviation of Normal to Normal peak intervals (SDNN) as heart rate variability (HRV) measures, which is a novel approach in studies measuring TOR-induced workload. Despite the study design that involved performing an $N = 1$ driving simulation experiment, significant differences between attribute levels have been found. This study provides recommendations on an empirically-validated set of design variables for future studies involving TORs and driver workload, specifically for the future study on personality and automated driving.

Keywords: Automated driving, Driving performance, Driving simulation, Raw Task Load Index, Heart rate, Heart rate variability, Validation, Workload

1. Introduction

At present, Level 2 automated vehicles ([1]) can be found on the road, where the driver has to monitor the Automated Driving System (ADS). Level 3 automated vehicles that allow the user to engage in non-driving related activities during automated driving are expected to find their way onto the road in the coming years [2]. However, the user must take-over the driving task in situations that exceed the Operational Design Domain (ODD) of the ADS. The take-over from automated to manual driving affects the safety of a level 3 automated vehicle and is crucial for successful implementation [3].

By developing meaningful human control for automated vehicles, the transition to Level 3 automated vehicles can be achieved responsibly [4]. A uniform approach may not be suitable to get drivers to take over the driving task from the ADS safely [5]. Possibly, individual differences in manual and automated driving can be included in the design of the ADS [4]. Including personality in automated driving is in line with the emergence of tailor-made Human-Machine-Interaction (HMI), which takes into account the complexity of the driving environment and driver's state for transferring control between the driver and the ADS [6]. It is already known that a link exists between personality and manual driving behaviour [7]. For instance, nervousness and anxiousness are attributed to Neuroticism (one of the Big Five [8], which is linked to a low tendency of risk-taking traffic behaviour [9]. However, it is unknown how personality is expressed in driver workload and thereby, automated driving behaviour and take-over performance.

Driver workload plays an important role in take-over performance, where both underload and overload can be detrimental [10]. Vehicle automation can result in mental underload when task demand is low, for instance during automated driving on the highway, or mental overload when task demand is high, for instance at the take-over [11, 12]. Workload depends on both context-dependent factors, such as traffic density at the take-over [13, 14] or engagement in a non-driving related task during automated driving [15, 16], and person-dependent factors, such as age and driving experience [17]. Possibly, personality plays a role in driver workload.

1.1. Aim of this research

Originally, this study aimed to investigate the role of personality in driver workload at take-overs by conducting a driving simulation experiment. However, due to the COVID-19 circumstances, the aim shifted to validating the design and research set-up of the aforementioned driving simulation experiment. This study will provide an empirically-validated set of design

*Corresponding author. *Email address:* `themis.marfoglia@gmail.com`

variables affecting driver workload for the study regarding the role of personality traits in automated driving.

This research is structured as follows. First, the methodology of the performed $N = 1$ self-experiment is presented, this includes the experimental apparatus, design of the simulation and the procedure of the experiment, as well as the used driver workload measures and statistical analyses. Then the results of the experiment are presented on the effect of the varying designs on driver workload. The findings are discussed, and the most suitable designs for the future study regarding personality are proposed. Finally, concluding remarks on the study are given.

2. Methods

An $N = 1$ experiment was performed with the researcher as the sole participant. Four design attributes are varied in the experiment, which are the time budget of the take-over request (TOR), traffic density at the take-over location, location of the TOR, and task during automated driving. Based on previous studies, these are expected to have a significant effect on driver workload [13, 14, 18, 19, 20, 21]. Various workload measures are included for measuring driver workload, including subjective and physiological measures, as well as driving performance indicators.

2.1. Apparatus

A driving simulator, located at the Department of Transport and Planning at Delft University of Technology was used for this study. The simulator is fitted with three high-resolution screens simulating the windshield and side window on both sides, providing a 180° field of view. As depicted in Figure 1, the three screens simulate the two side windows and the windshield of the vehicle. Furthermore, the simulator is equipped with a Fanatec haptic steering wheel and clutch, brake and gas pedal, a gear stick, hand brake, car seat and seat belt. Simulator data was logged at 50Hz. A tablet was placed on a holder on the right side of the driver's seat, used for playing Tetris as non-driving related task (NDRT) during automated driving.

Figure 1: Driving simulator used for the experiment

To measure heart rate (HR) and heart rate variability (HRV), an optical sensor was mounted on the participant's right index finger that measured light transmission through the fingertip (Figure 2). A photoplethysmography (PPG) was obtained, which was logged at 100Hz. An Atmel ATMega328p embedded processor board powers the recording of the data.

Figure 2: Finger sensor used for the experiment

Traditionally, the electrocardiogram (ECG) is the standard for measuring heart activity [22]. However, it was chosen to use PPG as it is a low-cost measure, and it measures heart activity in a less intrusive manner than ECG, which requires placing multiple electrodes on the participants' chest. PPG is, therefore, also less time-consuming compared to ECG. A disadvantage of PPG over ECG is, however, that it is sensitive to recording ambient noise [23]. By using HeartPy, an algorithm to handle heart activity data from PPG, developed by Van Gent et al. [23], the data can be filtered, and estimates for HR and HRV can be obtained, making PPG a valid alternative for ECG. A low-pass Butterworth filter with a cutoff of 3Hz is used for noise filtering.

Although HeartPy can filter noisy PPG data, large movements to the sensor may result in data collection problems which HeartPy cannot solve. For this, the Polar H10 chest strap was used as an additional method to measure HR (Figure 3). Polar H10 data was logged at 1Hz, which allows chest strap data to only be used as a substitute for mean HR data when data collection issues have occurred with the finger sensor. Chest strap data was used when the finger sensor data deviated more than 3 bpm (after the data filtering) from the mean HR measured by the chest strap [24].

Figure 3: Polar H10 chest strap used in the experiment [25]

To measure subjective workload, the Raw Task Load Index (RTLX) was used, which is a simplification of the NASA Task Load Index (TLX). By not requiring pairwise comparisons between the scales, it is a less time-consuming measure than the TLX [26, 27].

2.2. *Environment*

A 2x2 lane highway was simulated, with curves, a viaduct, and on- and off-ramps. Halfway the simulation, the highway made a large loop, after which it crossed the highway again via the viaduct. The simulation measures 7 km or 5.5 km for scenarios with a TOR on a straight road section or curve, respectively.

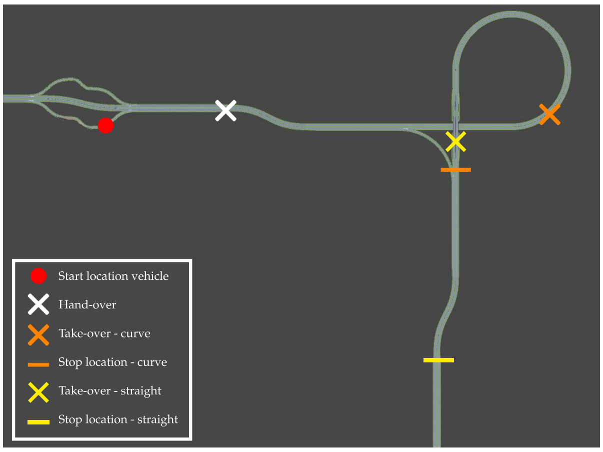


Figure 4: Highway layout of the driving simulation experiment

Trees and buildings are located near the highway in order to simulate a realistic highway environment. Figure 5 provides an impression of the scenario. During manual driving, the participant was allowed to choose their preferred speed, even though the participant was asked to adhere to the speed limit of 120 km/h.

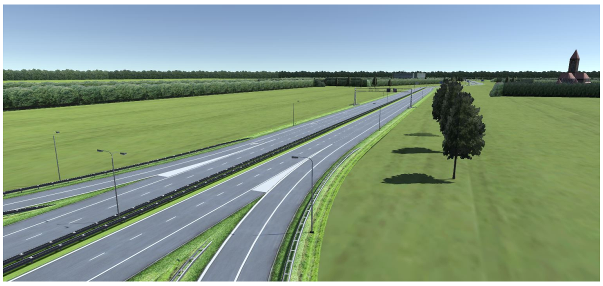


Figure 5: View on the highway

To vary the demand of the TOR, four design variables were varied in the experiment, which includes four time budgets (0, 5, 10, and 15 seconds), four traffic densities (zero: 0, low: 5, medium: 10, and high: 15 vehicles/km/lane), two TOR locations (straight stretch, and curve), and two tasks during automated driving (monitoring the simulated Automated Driving System (ADS) and vehicle, and Tetris as NDRT). In total, the experiment included 64 (= $4 * 4 * 2 * 2$) scenarios, all unique in their characteristics.

2.3. *Procedure*

Experiment days were limited between 9 AM and 5 PM and included a 1.5-hour lunch break and 10-15 minute breaks between sessions. The sessions included performing consecutive runs with a short time in between for setting up the next experiment and checking for correct placement of the finger sensor and chest strap. The participant played Tetris on the tablet during automated driving as NDRT, or was monitoring the simulated ADS and vehicle. After the run, the RTLX was filled in by the participant to assess subjective workload.

At the start of the experiment, the vehicle was located on the on-ramp. The participant was asked to merge onto the highway when feeling ready. After approximately 30 seconds of manual driving, a hand-over took place which shifts manual to automated driving. The driver would then proceed by monitoring the simulated ADS and vehicle or plays Tetris until the TOR. During automated driving, the vehicle adhered to the maximum speed of 120 km/h if possible. In the case of slower traffic driving in front of the ego-vehicle, the simulated ADS adjusted its speed to maintain a following distance of at least 1 second (50 meters at a speed of 120 km/h). If possible, the vehicle overtook slow-moving traffic. Without an apparent reason, the driver was notified of the TOR by an auditory signal after 2:00 minutes or 2:30 minutes when issued in a curve or straight road section, respectively. The driver was asked to take-over as soon as safely possible. The duration of manual driving after the take-over equalled 1 minute, which is the same duration as used in the studies by [13, 28] and [29]. Traffic is arranged with a constant and equal distance between the vehicles of 200, 100 and 66.6 meters for a traffic density of, respectively, 5, 10 or 15 vehicles/km/lane.

2.4. *Driver workload measures*

The following workload measures were used for measuring the effect of the various designs on driver workload:

- Scores on the six RTLX scales (mental demand, physical demand, temporal demand, performance, effort, and frustration), which are averaged to obtain the subjective overall workload score [27].
- Heart rate (HR) [bpm], calculated over a 1 minute duration before and after the TOR.
- Heart rate variability (HRV) measured by the RMSSD [ms] and SDNN [ms], calculated over a 1 minute duration before and after the TOR.
- Take-over reaction time [s], calculated as the duration between the TOR and first acceleration [30].
- Standard deviation of speed [km/h], calculated over 1 minute of manual driving after the take-over. Three seconds before and after a lane change were excluded from the analysis.
- Standard Deviation of Lateral Position (SDLP) [cm], calculated over 1 minute of manual driving after the take-over. Three seconds before and after a lane change were excluded from the analysis.

2.5. Statistical analyses

Tests for significance were performed by the one-way ANOVA or the Kruskal Wallis test when the residuals were not normally distributed. Levene's test was performed to check for equal variances. Welch and Brown-Forsythe tests were used instead of the ANOVA when the assumption for equal variances was violated (when $p < 0.05$). Games Howell is used for post hoc testing. As the same data-set was analysed multiple times, a Bonferroni correction was applied, which lowered the significance level to $p < 0.0125$. By applying a Bonferroni correction, the significance level ($p < 0.05$) is divided by the number of statistical tests performed on the same data: four tests to analyse the effect of time budget, traffic density, location of the TOR, and task engagement on workload.

3. Results

In total, 72 runs have been performed for 63 scenarios, as one scenario was accidentally not tested (instead, a different scenario was tested twice) and eight scenarios have been rerun due to expected data collection issues after a face validation of the data. After the data filtering, the computed mean heart rate (HR) over the entire run of the finger sensor data was compared to the computed HR by the Polar H10 chest strap. Nine scenarios encountered severe noise which resulted in inaccurate photoplethysmography (PPG) data. For these scenarios, chest strap data is used as a substitute for the mean HR. No missing or incorrect data is reported for the Raw Task Load Index (RTLX), chest strap, or driving metrics.

On average, a run took 3:46 minutes to complete, adding up to a total of 4:31 hours of driving time to complete all 72 runs. On the first day of experimenting 37 runs have been completed, on the second day 27 runs were completed and on the last day eight runs were completed as incorrect finger sensor data was expected for eight scenarios. During the 72 runs, no accidents have occurred that would have led to early termination of the run. The results of the driving simulation experiment are presented in Table 1.

3.1. Time budget

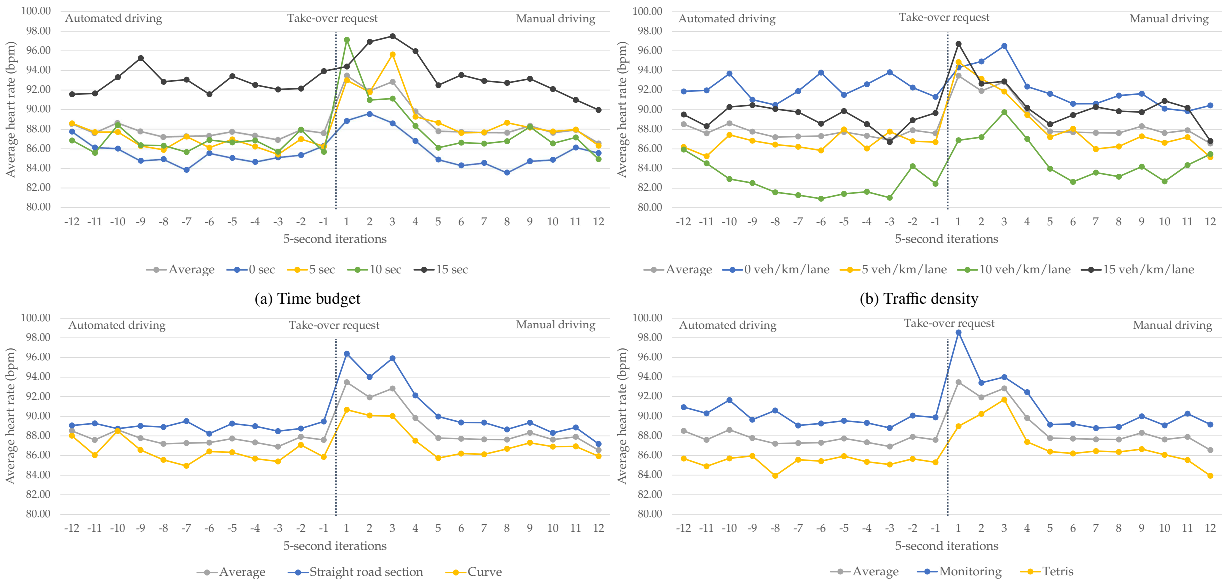
Workload as measured by the Raw Task Load Index (RTLX) decreased for every increase in time budget. Especially the 0-second time budget had a large effect on overall workload, with an average workload of 46.17%. As the assumption for equal variances did not hold ($F(3,59) = 7.725$, $p < 0.001$), the Welch and Brown-Forsythe tests were used and showed that the four time budgets were significantly different ($F(3,31.736) = 4.508$, $p = 0.010$ and $F(3,40.172) = 6.492$, $p = 0.001$, respectively). A post hoc analysis revealed a significant a significant higher workload in scenarios with a 0-second time budget compared to scenarios with a 15-second time budget ($p = 0.008$). Figure 6a illustrates the HR as it fluctuated 1 minute before and after the TOR. The figure shows an increased HR in scenarios with a 15-second time budget. HR increased immediately after the TOR, steeply in scenarios with 5- and 10-second time budgets, and less so for 0- and 15-second time budgets. Notably, scenarios with a 5-second time budget show a second peak at timestamp 3.

Despite the illustrated differences in HR response to the TOR for the different time budgets, the differences were found non-significant (ANOVA: $F(3,50) = 0.768$, $p = 0.517$). In addition, the response measured by the Root Mean Square of Successive Differences (RMSSD) (ANOVA: $F(3,50) = 0.284$, $p = 0.837$), and Standard Deviation of Normal to Normal peak intervals (SDNN) (Kruskal Wallis: $H(3) = 1.746$, $p = 0.627$) were found to be non-significant as well (Table 1). Moreover, although the illustrated mean HR differs between the various time budgets, no significant differences were found (ANOVA:

Table 1: The effect of the take-over request (TOR) on workload

	Subjective workload		**Physiological workload**				**Driving performance**			
		RTLX [%]		**HR [bpm]**	**RMSSD [ms]**	**SDNN [ms]**		**TOrt [s]**	**SD speed [km/h]**	**SDLP [cm]**
	N(*)	Mean(*SD*)	N(*)	Mean(*SD*)	Mean(*SD*)	Mean(*SD*)	N(*)	Mean(*SD*)	Mean(*SD*)	Mean(*SD*)
Average	64(1)	31.29(*17.66*)	64(10)	+ 1.02(*2.72*)	+ 0.31(*11.93*)	− 2.26(*12.97*)	64(1)	5.42(*2.14*)	5.19(*1.93*)	25.59(*10.88*)
Time budget [s]										
0	16(1)	46.17(*22.70*)[d]	16(1)	+ 0.33(*2.87*)	+ 1.20(*14.75*)	− 6.33(*15.47*)	16(1)	2.96(*1.80*)[bcd]	5.17(*2.02*)	31.00(*15.84*)
5	16(0)	30.31(*10.96*)	16(2)	+ 1.84(*2.88*)	− 0.53(*11.69*)	− 0.49(*12.92*)	16(0)	5.66(*1.14*)[a]	5.27(*2.17*)	25.19(*8.35*)
10	16(0)	27.39(*14.94*)	16(2)	+ 1.11(*2.76*)	+ 2.05(*13.62*)	+ 0.70(*11.10*)	16(0)	6.13(*1.70*)[a]	5.27(*1.91*)	22.31(*8.71*)
15	16(0)	22.24(*11.82*)[a]	16(5)	+ 0.74(*2.27*)	− 2.06(*5.42*)	− 2.73(*11.72*)	16(0)	6.81(*1.71*)[a]	5.04(*1.79*)	24.19(*8.17*)
Traffic density [vehicles/km/lane]										
0	16(0)	24.69(*18.58*)	16(2)	− 0.26 (*2.84*)	− 4.36(*6.86*)	− 2.00(*11.66*)	16(0)	5.30(*1.90*)	4.12(*2.05*)	16.21(*5.12*)[bcd]
5	16(0)	34.01(*16.53*)	16(3)	+ 1.13(*2.20*)	+ 0.39(*11.56*)	− 2.74(*7.28*)	16(0)	4.66(*2.00*)	4.78(*2.28*)	29.92(*14.25*)[a]
10	16(0)	33.80(*19.07*)	16(3)	+ 2.48(*2.59*)	− 0.53(*18.22*)	− 3.45(*21.90*)	16(0)	5.39(*2.34*)	5.77(*1.70*)	23.54(*6.90*)[a]
15	16(1)	32.78(*16.18*)	16(1)	+ 0.84(*2.72*)	+ 1.73(*10.01*)	− 0.97(*7.40*)	16(1)	5.78(*2.61*)	6.12(*1.72*)	32.07(*9.64*)[a]
Location TOR										
Straight	32(1)	21.99(*11.85*)[b]	32(4)	+ 1.15(*2.45*)	+ 0.73(*8.71*)	− 0.29(*11.09*)	32(1)	5.41(*2.32*)	4.41(*1.83*)[b]	23.87(*11.00*)
Curve	32(0)	40.31(*17.82*)[a]	32(5)	+ 0.89(*3.00*)	− 0.11(*14.63*)	− 4.23(*14.56*)	32(0)	5.45(*1.98*)	5.93(*1.75*)[a]	27.25(*10.67*)
Task during automated driving										
Monitoring	32(0)	30.62(*19.47*)	32(5)	+ 0.52(*2.94*)	+ 0.53(*9.92*)	− 5.82(*14.37*)	32(0)	5.07(*2.27*)	5.18(*2.00*)	25.63(*12.46*)
Tetris	32(1)	31.99(*15.88*)	32(4)	+ 1.52(*2.42*)	+ 0.09(*13.85*)	+ 1.30(*10.49*)	32(1)	5.80(*1.96*)	5.20(*1.89*)	25.55(*9.18*)

* Number of excluded runs because of data collection errors.
[abcd] Significant difference at the 1.25% level with first ([a]), second ([b]), third ([c]), fourth ([d]) attribute level.



(a) Time budget (b) Traffic density

(c) Location TOR (d) Task during automated driving

Figure 6: HR fluctuation in in 5-second timestamps of a minute measurement before and after the TOR in different task load conditions

$F(3, 59) = 0.304$, $p = 0.823$). The differences in mean RMSSD (ANOVA: $F(3, 50) = 1.040$, $p = 0.124$) and SDNN (ANOVA: $F(3, 50) = 1.566$, $p = 0.209$) were non-significant as well (Table 2).

Table 2: Mean physiological workload, measured over the entire run

	HR [bpm]		RMSSD [ms]		SDNN [ms]
	N(*)	Mean(*SD*)	N(*)	Mean(*SD*)	Mean(*SD*)
Average	64(1)	88.19(*7.28*)	64(10)	36.91(*13.95*)	36.87(*12.97*)
Time budget [s]					
0	16(1)	87.64(*9.42*)	16(1)	39.92(*18.13*)	40.46(*15.83*)
5	16(0)	88.49(*5.84*)	16(2)	37.06(*14.67*)	35.80(*9.63*)
10	16(0)	87.14(*5.58*)	16(2)	38.48(*8.08*)	38.57(*8.84*)
15	16(0)	89.47(*8.21*)	16(5)	30.62(*12.11*)	31.20(*8.33*)
Traffic density [veh/km]					
0	16(0)	92.25(*7.91*)	16(2)	29.18(*9.64*)[c]	31.16(*7.45*)[c]
5	16(0)	87.20(*5.47*)	16(3)	36.25(*10.80*)	36.74(*10.47*)
10	16(0)	84.74(*4.93*)	16(3)	48.28(*16.95*)[a]	45.40(*13.62*)[a]
15	16(1)	88.61(*8.69*)	16(1)	34.69(*11.34*)	34.80(*10.00*)
Location take-over request (TOR)					
Straight	32(1)	89.74(*7.95*)	32(4)	34.07(*12.79*)	34.41(*10.05*)
Curved	32(0)	86.69(*6.33*)	32(5)	39.76(*14.72*)	39.34(*12.51*)
Task during automated driving					
Monitoring	32(0)	89.80(*8.50*)	32(5)	34.86(*15.68*)	36.41(*13.26*)
Game	32(1)	86.53(*5.39*)	32(4)	38.97(*11.93*)	37.34(*9.69*)

* Number of excluded runs because of data collection errors.
[abcd] Significant difference at the 1.25% level with first ([a]), second ([b]), third ([c]), fourth ([d]) attribute level.

The mean take-over reaction time (TOrt) differed significantly between the time budgets (Kruskal Wallis: $H(3) = 27.440$, $p < 0.001$). The 0-second time budget resulted in a significantly lower TOrt compared to the 5-second time budget ($p = 0.003$), to the 10-second time budget ($p = 0.001$), and to the 15-second time budget ($p < 0.001$). The time budget was exceeded in 25 scenarios, evidently this comprised of all 0-second time budget scenarios, and also 10 scenarios with a 5-second time budget. Thus, the exceedance in 0-second time budget scenarios equalled the TOrt reported in Table 1 of M(*SD*) = 2.96(*1.80*) seconds. The exceedance in scenarios with a 5-second time budget equalled M(*SD*) = 1.23(*0.90*) seconds. The duration of the exceeded time budget is significantly longer in scenarios with a 0-second time budget (Kruskal-Wallis $H(1) = 10.351$, $p = 0.001$).

The driver's longitudinal control ability as represented by the standard deviation (SD) of driver speed does show small differences between the time budgets which were non-significant (ANOVA: $F(3, 59) = 0.050$, $p = 0.985$). No significant differences were also found for lateral control ability (ANOVA: $F(3, 59) = 1.894$, $p = 0.140$).

3.2. *Traffic density*

The tabulated differences in subjective overall workload suggest that scenarios without traffic resulted in less workload (Table 1). However, these differences were found non-significant (ANOVA: $F(3, 59) = 1.016$, $p = 0.392$).

For physiological workload, the lowest mean HR of 84.74 bpm was measured in scenarios with medium traffic density (10 vehicles/km/lane), whereas the highest mean HR of 92.25 bpm was measured in scenarios without any traffic (Table 2 and Figure 6b). Despite this illustrated and tabulated difference, these differences were found non-significantly different (ANOVA, $F(3, 50) = 3.311$, $p = 0.026$; Bonferroni correction). With increasing traffic density, the mean RMSSD and SDNN also increased. However, this is not the case for high traffic density, which resulted in the second lowest mean RMSSD and

SDNN (Table 2). The differences in RMSSD were found significant (ANOVA: $F(3, 50) = 5.589$, $p = 0.002$), as well as the differences in SDNN (ANOVA: $F(3, 50) = 5.589$, $p = 0.008$). Both RMSSD and SDNN reported a significant difference between zero ($p = 0.001$) and medium traffic density scenarios ($p = 0.005$).

An increase in HR after the TOR is reported for all traffic densities, with an immediate peak after the TOR for low and high traffic density, whereas zero and medium traffic density showed a delayed peak at timestamp 3. The size of the HR-response increased with increases in traffic density, except for high traffic density that resulted in the second highest mean HR (Table 2). However, again, the found differences in HR response were non-significant (ANOVA: $F(3, 50) = 2.518$, $p = 0.069$). The TOR response measured by the RMSSD and SDNN differs. The RMSSD showed an irregular increase and decrease in RMSSD, whereas the SDNN always decreased after the TOR. The differences in RMSSD were non-significant (ANOVA: $F(3, 50) = 0.099$, $p = 0.960$). Although the size of the SDNN response increased when traffic density increased from zero to medium traffic density (the smallest effect size is reported in high traffic density scenarios: Table 2), however the differences were found non-significant (Kruskal Wallis: $H(3) = 0.598$,$p = 0.897$).

The TOrt increased from low to medium traffic density. The second shortest TOrt was reported for zero traffic density scenarios (Table 1). However, no significant differences in TOrt for the different traffic densities were found (Kruskal Wallis: $H(3) = 0.873$, $p = 0.832$). The available time budget was exceeded in 7 scenarios with zero traffic density (M(*SD*) = 3.06(*2.62*)), and in 6 scenarios with low (M(*SD*) = 1.90(*0.91*)), medium (M(*SD*) = 2.46(*1.53*)), and high traffic density (M(*SD*) = 1.54(*1.03*)). The differences in duration of the exceedance is found non-significant (Kruskal Wallis: $H(3) = 2.241$, $p = 0.524$).

Speed deviation increased for every increase in traffic density (Table 1), but the differences were found non-significant(ANOVA: $F(3, 50) = 3.201$, $p = 0.030$; Bonferroni correction). Moreover, the mean Standard Deviation of Lateral Position (SDLP) for the different traffic densities increased with increases in traffic density, except for medium traffic density scenarios for which the second lowest SDLP was reported. A significant difference was found between the mean SDLP in different traffic densities (ANOVA: $F(3, 62) = 8.193$, $p < 0.001$). The SDLP in scenarios without traffic was significantly smaller compared to the SDLP found in low ($p = 0.007$), medium ($p = 0.002$), and high ($p < 0.001$) traffic density.

3.3. Location of the take-over request

Higher subjective overall workload was reported in scenarios with a TOR in a curve: 40.31% vs 21.99% on a straight road stretch. The Welch and Brown-Forsythe tests showed that this workload difference was significant ($F(1, 54.100) = 23.228$, $p < 0.001$) in both tests.

A similar pattern in HR during the experiment is reported for both TOR locations (Figure 6c). A higher mean HR was reported in scenarios with the TOR on a straight road section, the difference however was found non-significant (ANOVA: $F(1, 61) = 2.850$, $p = 0.096$). For scenarios with a TOR in a curve, a higher RMSSD and SDNN was reported. However, the differences were found non-significant for the RMSSD (ANOVA: $F(1, 52) = 2.301$, $p = 0.135$) and SDNN (ANOVA: $F(1, 52) = 2.542$, $p = 0.117$).

The TOR resulted in two HR peaks in scenarios with a TOR on a straight road section (timestamp 1 and 3), whereas the HR after the TOR in a curve increased immediately in timestamp 1 and remained stable until timestamp 3, after which it decreased to the pre-TOR HR. An increase in HR is reported for both TOR locations, but the difference between the two was found non-significant (ANOVA: $F(1, 52) = 0.124$, $p = 0.726$). Regarding the effect of the TOR on RMSSD, an increase of +0.73 ms was reported in straight road section scenarios, whereas a decrease of -0.11 ms is reported in curve scenarios. Using the Welch and Brown-Forsythe tests it was found that the difference was non-significant ($F(1, 42.394 = 0.065$, $p = 0.799$) in both tests. The SDNN decreased after the TOR for both locations of the TOR. The difference in size of the effect was found non-significant (Kruskal Wallis: $H(1) = 0.399$, $p = 0.528$).

The TOrt for both take-over locations showed small differences, which were non-significant (Kruskal Wallis: $H(1) = 0.048$ $p = 0.826$). The distribution of exceeded time budgets was also almost equal between the two locations: 13 out of 31 scenarios with a TOR on a straight road section (M(*SD*) = 1.62(*1.02*)) and 12 out of 32 scenarios with a TOR in a curve (M(*SD*) = 2.98(*1.02*)). The difference in exceedance between the two TOR locations was found not to differ (Kruskal Wallis: $H(1) = 3.834$, $p = 0.050$).

Speed deviation was significantly higher in scenarios with a TOR in a curve (ANOVA: $F(1, 61) = 11.394$, $p = 0.001$). The results tabulated in Table 1 suggest that SDLP was greater in scenarios with a TOR in a curve, although the difference was found non-significant (ANOVA: $F(1, 61) = 1.531$, $p = 0.221$).

3.4. Non-driving related task engagement

Overall workload as measured by the RTLX reported a slightly higher workload when Tetris was played during automated driving (30.62% vs 31.99% when monitoring). The difference, however, was found non-significant (ANOVA: $F(1, 61) = 0.093$, $p = 0.762$).

A higher HR during the experiment was reported in monitoring scenarios (89.80 bpm vs. 86.53 bpm). However, this difference was found non-significant (ANOVA: $F(1, 61) = 3.287$, $p = 0.075$). The RMSSD and SDNN after the TOR were higher in scenarios when Tetris was played during automated driving. However, no significant difference was found in mean RMSSD (ANOVA: $F(1, 52) = 1.173$, $p = 0.284$) and SDNN (ANOVA: $F(1, 52) = 0.086$, $p = 0.771$) between the tasks.

HR during the experiment increased steeply immediately after the TOR in scenarios with a monitoring task, whereas Tetris scenarios showed a delayed peak at timestamp 3. On average, an increase in HR is reported for both tasks after resumption of manual control: + 1.52 bpm when monitoring and + 0.52 when playing Tetris. This difference was found non-significant

(ANOVA: $F(1, 52) = 1.880$, $p = 0.176$). The RMSSD also reported an increase after the TOR for both tasks, again, the difference in effect size was found non-significant (ANOVA: $F(1, 52) = 0.892$, $p = 0.892$). In contrast, the SDNN reported a decrease in SDNN after the TOR in monitoring scenarios, whereas an increase is reported when Tetris was played. However, again, the difference in effect size was found non-significant (Kruskal Wallis: $H(1) = 3.023$, $p = 0.082$).

The TOrt was longer in scenarios with Tetris as non-driving related task (NDRT) (Table 1), although no significant difference was found between the TOrts of the two tasks (Kruskal Wallis: $H(1) = 2.084$, $p = 0.149$). The time budget was exceeded 13 times in scenarios when the driver was monitoring with a duration of M(*SD*) = 1.65(*0.82*) seconds, and 12 times when Tetris was played with a duration of M(*SD*) = 2.95(*2.18*) seconds. The duration of the exceedance was significantly higher with Tetris as NDRT (Kruskal Wallis: $H(1) = 4.734$, $p = 0.030$).

Longitudinal control ability as represented by the SD speed did not differ between the tasks (Table 1) as indicated by a non-significant finding (ANOVA: $F(1, 61) = 0.002$, $p = 0.968$). Lateral control ability, as indicated by the SDLP was also similar for the two tasks, as indicated by the non-significant difference (ANOVA: $F(1, 61) = 0.001$, $p = 0.978$).

4. Discussion

Although design variables and respective attributes were chosen, which were expected to result in significant workload differences, this was only found on a limited scale. This has two reasons. Firstly, because of the $N = 1$ study design, the number of observations was limited. Secondly, the fact that the number of variables to be analysed in this study was maximised in as few runs as possible, resulting in large standard deviations (SDs) in the workload measurements. It is therefore already exceptional that significant differences were found. The workload differences found between the various designs provide sufficient indication of finding significant workload differences in the future experiment regarding personality (with >100 participants). This indication is reinforced when significant differences were demonstrated in the current $N = 1$ study.

4.1. Take-over requests and their effect on workload

All scenarios combined, an overall workload as measured by the Raw Task Load Index (RTLX) of 31.29% was found, which lies within the expected range for subjective workload measured by the RTLX in simulated automated driving studies [31]. Specifically, [32], who also required a required a regular response to take-over requests (TORs) (i.e. every three minutes), measured a similar overall workload of 31%.

For physiological workload, a mean heart rate (HR) of 88.19 bpm was measured during the experiment, which lies within one SD of the mean resting HR of M(*SD*) = 80.2(14.8) [33]. The measured Root Mean Square of Successive Differences (RMSSD) of 36.91 ms and Standard Deviation of Normal to Normal peak intervals (SDNN) of 36.87 ms fall within the standards for short-term measurements of M(*SD*) = 42(15) ms for RMSSD and M(*SD*) = 50(16) ms for SDNN [34]. Simulated automated driving was therefore not experienced as very demanding, but it could also not be considered an easy or relaxing activity.

The average increase of +1.02 bpm in HR and decrease of −2.26 ms in SDNN after the TOR indicate that the TOR increased workload. In contrast, the average increase of +0.31 ms in RMSSD after the TOR indicates a reduced workload after the TOR. [35] and [36] also measured the difference in HR after a TOR and reported an increase in HR of, respectively, +0.43 bpm and +2.98 bpm after the TOR. The results of this study are thus within the expected range, although based on [36], it could be argued that these results indicate a small effect size.

This study included a novel approach using the RMSSD and SDNN to measure TOR-induced workload. Currently, there is no consensus on the usefulness of these heart rate variability (HRV) measures as workload measure [37, 38, 39]. As far as is known, only [40] (published at the time of writing this study) used RMSSD and SDNN to measure TOR-induced workload. [40] conducted an exploratory study with two participants who experienced four TORs under two weather conditions (sunny / rain) and alert modalities (visual-auditory / auditory). However, their data cannot be translated to this study, as [40] only presented the mean RMSSD and SDNN after the TOR instead of the respective increases or decreases after the TOR. Other studies used the RMSSD and SDNN to measure workload differences between low and high task demands during automated driving [37, 38, 39, 41, 42, 43]. Based on the results of this study, the SDNN appears to be a more sensitive workload measure than RMSSD, as RMSSD often indicated a different direction of the expected effect on workload for the various design variables. SDNN appeared sensitive for distinguishing the highest task demands, but less sensitive for distinguishing between low and medium task demands. Future studies using the RMSSD and SDNN to measure specifically TOR-induced workload are necessary to gain a better understanding of the sensitivity of RMSSD and SDNN as (TOR-induced) workload measure, compared to other proven physiological workload measures such as HR.

This study found a relatively high take-over reaction time (TOrt) of 5.42 seconds, which is considerably longer than the TOrt found by [18] of 2.06 and 3.10 seconds for 5- and 7-second time budgets, respectively. The used design of the take-over attributed to this increased TOrt. Usually, a button on the steering wheel or pressing a driving pedal is used to resume manual driving before the expiration of the time budget. However, this study required the participant to press a key combination on the keyboard to resume driving before the expiration of the available time budget. As the keyboard was placed on the dashboard of the simulator, more time was needed to reposition to be able to press the key-combination, which resulted in increased TOrts. Therefore, for future experiments it is recommended to use a button on the steering wheel to take-over the driving task, or by allowing overruling of the system when the steer, gas pedal, or brake is used. The other driving performance measures reported values within the expected range, with an average speed variation of 5.19 km/h and Standard Deviation of Lateral Position (SDLP) of 25.59 cm. For example, [44] found a speed deviation of 5.9 km/h, and [45] reported an average SDLP between 15 and 30 cm after a take-over at a speed of 50 km/h.

4.2. Time budget as design variable

The various workload measures did not unambiguously indicate differences in TOR-induced workload with different time budgets. It was expected that every increase in time budget would result in a smaller effect of the TOR on workload [46]. The RTLX suggests, indeed, that workload decreases with a longer available duration to respond to the TOR. Similarly, the SDLP suggests that the 0-second time budget has a greater effect on workload than time budgets of 5 seconds or greater. However, driving performance, as indicated by the SDLP, did not seem to improve with a time budget duration greater than 5 seconds. The SDNN showed how immediate disengagement of the Automated Driving System (ADS) resulted in greater workload compared to scenarios with a time budget of 5 seconds or greater. So if an urgent take-over is to be simulated, it is recommended to use a 0-second time budget instead of a 5-second time budget. Moreover, time budgets of 15 seconds are recommended for non-urgent take-overs, as the RTLX found a significant difference between 0- and 15-second time budgets. Although, it could also be argued that a 10-second time budget could be used for non-urgent take-overs. Namely, already in this $N = 1$ study a significance level of $p = 0.056$ was found for the difference between the 0- and 10-second time budget.

4.3. Varying traffic densities and their effect on workload

It has been found that driving performance after the TOR decreased with increases in traffic density in the current experimental set-up. Possibly, overload developed due to increases in traffic density, which is detrimental for driving performance [47]. However, the subjective and physiological workload measures did not fully indicate overload in high traffic density scenarios. Namely, less workload was measured in scenarios with high traffic density by the RTLX, HR, and SDNN. So, if in a future experiment, a low demand take-over is simulated, it is preferred to use zero traffic density over a low traffic density of 5 vehicles/km/lane, as the low and medium traffic density induced similar workload. If high task demand must be varied, it is recommended to simulate a low or medium traffic density of 5 or 10 vehicles per kilometre per lane. However, preference is given to simulate a medium traffic density, as already in this experimental set-up, significant workload differences were found between no-traffic and medium traffic density scenarios. Simulating a high traffic density of 15 vehicles per kilometre per lane was not ideal in the current design of the simulation. The ego-vehicle remained on the right lane during automated driving. At the take-over, a lower initial speed is found compared to the other traffic densities, which could have resulted in less TOR-induced workload in high traffic density scenarios.

4.4. The location of the take-over request

In the current experiment set-up, the TOR in a curve induced more workload compared to a TOR on a straight road section. The difference in workload was most evident in the difference in subjective workload measured by the RTLX. Physiologically, the difference in workload is also found by the RMSSD and SDNN. Post-take-over driving performance is noticeably worse when the TOR was issued in a curve, as measured by the speed deviation and SDLP. However, a distorted TOrt is possibly found; the mean TOrt in scenarios with a TOR in a curve is possibly shorter than measured. The used TOrt calculation method could have attributed to the increased TOrt, which measured the time between the TOR and the first acceleration. As the speed of the vehicle at the take-over in a curve is approximately 120 km/h, the driver delays acceleration to reach a desired lower speed for driving in a curve. Therefore, the measured TOrt is higher than the actual TOrt, as the driver was already using the steering wheel for lane keeping. Therefore, it is recommended to use a different calculation method for the TOrt based on the steering angle. For future experiments, the use of both locations is encouraged for distinguishing between low and high task demand, as evident workload differences were measured.

4.5. Engagement in a non-driving related task

It was expected that a clear difference in the effect of the TOR on workload could be distinguished between the two tasks, i.e. when engaged in Tetris, the TOR was expected to have a greater effect on workload compared to the monitoring task. However, the results suggest that monitoring led to underload, resulting in a greater effect of the TOR on workload. Moreover, the results also suggest that playing Tetris did not resulted in underload, as was expected based on [47]. In the current experimental set-up, no clear differences in the effect of workload was measured between the two tasks. Possibly, if the duration of automated driving before the TOR was extended, playing Tetris would have resulted in underload, and therefore the TOR would have resulted in greater workload increases. However, due to the short duration of automated driving, playing Tetris did not result in drowsiness, rather it resulted in increased driver vigilance.

5. Conclusions

This study aimed to investigate the role of personality in drivers' workload induced by take-over requests (TORs) by a driving simulation experiment. To ultimately make recommendations on how to incorporate personality into the design of Automated Driving Systems (ADSs), and by making recommendations on tailor-taught behaviour in automated vehicles. However, due to the COVID-19 conditions, the aim of this study changed to validating the design of the driving simulation study using an $N = 1$ approach. With this new aim, recommendations were also provided for future $N = 1$ studies.

Four design characteristics were varied that, based on literature, were expected to have a significant effect on workload, i.e. time budgets of the TOR, traffic densities, the location of the TOR, and engagement in a non-driving related task (NDRT) during automated driving. Variations of these design characteristics were tested in the simulation to determine which variation is best suited to use in the experiment. In terms of time budget, it is recommended to use 0 seconds for urgent (high task demand) take-overs and 10 or 15 seconds for non-urgent (low task demand) take-overs. For traffic density, it is recommended to simulate no traffic for a take-over in low task demand. If traffic is to be simulated, less than 5 vehicles/km/lane is recommended, as this traffic density increased workload considerably. For a high task demand take-over, it is recommended to either simulate 10 or 15 vehicles/km/lane. Although in the current experiment set-up, physiological indicators did not reflect the greatest workload for high traffic density, it is expected that this will be the case in an experiment that assures an equal speed during automated driving. Regarding the location of the TOR, issuing is recommended on a straight stretch for low demand take-overs and in a curve for high demand take-overs. For simulating underload as a result of NDRT engagement, it is advised to simulate a longer duration of NDRT engagement than used in this study. With the brevity of NDRT engagement in this study, instead of underload, increased driver vigilance was simulated. In light of the COVID-19 conditions that complicated conducting studies involving participants, conducting an $N = 1$ study proved to be a valuable research method for validating the design of the driving simulation experiment. Besides recommendations to improve the design of the driving simulation experiment, the study provided an indication of the usefulness of including the Standard Deviation of Normal to Normal peak intervals (SDNN) as physiological workload measure. In the absence of the need of including participants other than the researcher, the $N = 1$ study proved to be a more accessible and easily applicable method of validating research set-ups. Based on the results of an $N = 1$ study, hypotheses can be formulated as to whether or not research results can be expected, which allows a more focused approach in future studies.

Acknowledgements

This study is funded through the NWO-sponsored project, called Meaningful Human Control over Automated Driving Systems [Project Number: 313-99-329]

Research data is available via 4TU, DOI: 10.4121/13102763.

References

[1] SAE International, Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles, Tech. rep., SAE International, Washington, DC (2018).

[2] European Commission, On the road to automated mobility: an EU strategy for mobility of the future, Tech. rep., European Commission (2018).
URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0283

[3] C. Gold, R. Happee, K. Bengler, Modeling take-over performance in level 3 conditionally automated vehicles, Accident Analysis and Prevention 116 (2018) 3–13. doi:10.1016/j.aap.2017.11.009.

[4] D. D. Heikoop, M. Hagenzieker, G. Mecacci, S. Calvert, F. S. D. Sio, B. v. Arem, Human behaviour with automated driving systems: a quantitative framework for meaningful human control, Theoretical Issues in Ergonomics Science (2019). doi:10.1080/1463922X.2019.1574931.

[5] M. Körber, K. Bengler, Potential individual differences regarding automation effects in automated driving, in: INTERACCIÓN 2014: Proceedings of the XV International Conference on Human Computer Interaction, Puerto de la Cruz, Tenerife, Spain, 2014, pp. 1–7. doi:10.1145/2662253.2662275.

[6] T. Dukic Willstrand, A. Anund, M. Pereira Cocron, S. Griesche, N. Strand, S. Troberg, L. Zanovello, Collecting end-users needs regarding driver state-based automation in the ADAS&ME project, in: Proceedings of 7th Transport Research Arena TRA 2018, Vienna, Austria, 2018, pp. 1–10.
URL https://www.adasandme.com/wp-content/uploads/2018/06/TRA2018_final_Wilstrand-II.pdf

[7] O. Taubman-Ben-Ari, M. Mikulincer, O. Gillath, The multidimensional driving style inventory - scale construct and validation, Accident Analysis and Prevention 36 (3) (2004) 323–332. doi:10.1016/S0001-4575(03)00010-1.

[8] R. Cattell, Personality and motivation structure and measurement, Harcourt, Brace, Jovanovich, New York, 1957.

[9] O. Taubman-Ben-Ari, D. Yehiel, Driving styles and their associations with personality and motivation, Accident Analysis and Prevention 45 (2012) 416–422. doi:10.1016/j.aap.2011.08.007.

[10] M. S. Young, N. A. Stanton, Attention and automation: new perspectives on mental underload and performance, Theoretical Issues in Ergonomics Science 3 (2) (2002) 178–194. doi:10.1080/14639220210123789.

[11] J. F. Coughlin, B. Reimer, B. Mehler, Driver wellness, safety & the development of an awarecar, Tech. rep., Massachusetts Institute of Technology, Cambridge, Massachusetts (2009).
URL http://web.mit.edu/reimer/www/pdfs/coughlin_wellness_2009.pdf

[12] J. Paxion, E. Galy, C. Berthelon, Mental workload and driving, Frontiers in Psychology 5 (1344) (2014) 1–11. doi:10.3389/fpsyg.2014.01344.

[13] C. Gold, M. Körber, D. Lechner, K. Bengler, Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density, Human factors 58 (4) (2016) 642–652. doi:10.1177/0018720816634226.

[14] M. S. L. Scharfe, K. Zeeb, N. Russwinkel, The Impact of Situational Complexity and Familiarity on Takeover Quality in Uncritical Highly Automated Driving Scenarios, Information 11 (115) (2020) 1–12. doi:10.3390/info11020115.

[15] N. Merat, A. H. Jamson, F. C. Lai, O. Carsten, Highly automated driving, secondary task performance, and driver state, Human Factors 54 (5) (2012) 762–771. doi:10.1177/0018720812442087.

[16] K. Zeeb, A. Buchner, M. Schrauf, Is take-over time all that matters? The impact of visual-cognitive load on driver take-over quality after conditionally automated driving, Accident Analysis and Prevention 92 (2016) 230–239. doi:10.1016/j.aap.2016.04.002.

[17] T. J. Wright, S. Samuel, A. Borowsky, S. Zilberstein, D. L. Fisher, Experienced drivers are quicker to achieve situation awareness than inexperienced drivers in situations of transfer of control within a Level 3 autonomous environment, in: Proceedings of the Human Factors and Ergonomics Society, Sage Publications, 2016, pp. 270–273. doi:10.1177/1541931213601062.

[18] C. Gold, D. Damböck, L. Lorenz, K. Bengler, Take over! How long does it take to get the driver back into the loop?, in: Proceedings of the Human Factors and Ergonomics Society, Vol. 57, 2013, pp. 1938–1942. doi:10.1177/1541931213571433.

[19] N. A. Stanton, M. S. Young, Driver behaviour with adaptive cruise control, Ergonomics 48 (10) (2005) 1294–1313. doi:10.1080/00140130500252990.

[20] B. Mok, M. Johns, K. J. Lee, D. Miller, D. Sirkin, P. Ive, W. Ju, Emergency, Automation Off: Unstructured Transition Timing for Distracted Drivers of Automated Vehicles, in: 2015 IEEE 18th International Conference on Intelligent Transportation Systems, IEEE, Las Palmas, 2015, pp. 2458–2464. doi:10.1109/ITSC.2015.396.

[21] F. Naujoks, D. Befelein, K. Wiedemann, A. Neukum, A review of non-driving-related tasks used in studies on automated driving, in: N. Stan-

ton (Ed.), Proceedings of the AHFE 2017 International Conference on Human Factors in Transportation, Springer, Cham, 2017, pp. 525–537. doi:10.1007/978-3-319-60441-1{_}52.
[22] N. Selvaraj, A. Jaryal, J. Santhosh, K. K. Deepak, S. Anand, Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography, Journal of Medical Engineering and Technology 32 (6) (2008) 479–484. doi:10.1080/03091900701781317.
[23] P. van Gent, H. Farah, N. van Nes, B. van Arem, Analysing Noisy Driver Physiology Real-Time Using Off-the-Shelf Sensors: Heart Rate Analysis Software from the Taking the Fast Lane Project, Journal of Open Research Software 7 (10 2019). doi:10.5334/jors.241.
[24] E. Jo, K. Lewis, D. Directo, M. J. Y. Kim, B. A. Dolezal, Validation of biofeedback wearables for photoplethysmographic heart rate tracking, Journal of Sports Science and Medicine 15 (3) (2016) 540–547. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4974868/
[25] Polar, Polar H10 (2020). URL https://www.polar.com/nl/producten/accessoires/h10_hartslagsensor
[26] S. G. Hart, L. E. Staveland, Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, Advances in Psychology 52 (C) (1988) 139–183. doi:10.1016/S0166-4115(08)62386-9.
[27] S. G. Hart, NASA-task load index (NASA-TLX); 20 years later, in: Proceedings of the Human Factors and Ergonomics Society, 2006, pp. 904–908. doi:10.1177/154193120605000909.
[28] M. Bueno, E. Dogan, F. H. Selem, E. Monacelli, S. Boverie, A. Guillaume, How different mental workload levels affect the take-over control after automated driving, in: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, IEEE, 2016, pp. 2040–2045. doi:10.1109/ITSC.2016.7795886.
[29] M. Körber, C. Gold, D. Lechner, K. Bengler, The influence of age on the take-over of vehicle control in highly automated driving, Transportation Research Part F: Traffic Psychology and Behaviour 39 (2016) 19–32. doi:10.1016/j.trf.2016.03.002.
[30] V. Melcher, S. Rauh, F. Diederichs, H. Widlroither, W. Bauer, Take-over requests for automated driving, Procedia Manufacturing 3 (2015) 2867–2873. doi:10.1016/j.promfg.2015.07.788.
[31] D. D. Heikoop, J. C. De Winter, B. Van Arem, N. A. Stanton, Acclimatizing to automation: Driver workload and stress during partially automated car following in real traffic, Transportation Research Part F: Traffic Psychology and Behaviour 65 (2019) 503–517. doi:10.1016/j.trf.2019.07.024.
[32] J. De Winter, N. A. Stanton, J. C. F. De Winter, N. A. Stanton, J. S. Price, H. Mistry, N. Stanton, The effects of driving with different levels of unreliable automation on self-reported workload and secondary task performance, International Journal of Vehicle Design 70 (4) (2016) 297:324. doi:10.1504/IJVD.2016.076736.
[33] R. Avram, G. H. Tison, K. Aschbacher, P. Kuhar, E. Vittinghoff, M. Butzner, R. Runge, N. Wu, M. J. Pletcher, G. M. Marcus, J. Olgin, Real-world heart rate norms in the Health eHeart study, NPJ Digital Medicine 2 (58) (2019) 1–10. doi:10.1038/s41746-019-0134-9.
[34] F. Shaffer, J. P. Ginsberg, An overview of heart rate variability metrics and norms, Frontiers in Public Health 5 (258) (2017) 1–17. doi:10.3389/fpubh.2017.00258.
[35] M. T. Alrefaie, S. Summerskill, T. W. Jackon, In a heart beat: Using driver's physiological changes to determine the quality of a takeover in highly automated vehicles, Accident Analysis and Prevention 131 (2019) 180–190. doi:10.1016/j.aap.2019.06.011.
[36] D. Ruscio, A. J. Bos, M. R. Ciceri, Distraction or cognitive overload? Using modulations of the autonomic nervous system to discriminate the possible negative effects of advanced assistance system, Accident Analysis and Prevention 103 (2017) 105–111. doi:10.1016/j.aap.2017.03.023.
[37] B. Mehler, B. Reimer, Y. Wang, A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload, in: Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, IA Public Policy Center, Olympic Valley, Lake Tahoe, California, Iowa City, 2011, pp. 590–597. doi:10.17077/drivingassessment.1451.
[38] A. Luque-Casado, J. C. Perales, D. Cárdenas, D. Sanabria, Heart rate variability and cognitive processing: The autonomic response to task demands, Biological Psychology 113 (2016) 83–90. doi:10.1016/j.biopsycho.2015.11.013.
[39] A. R. Hidalgo-Muñoz, A. J. Béquet, M. Astier-Juvenon, G. Pépin, A. Fort, C. Jallais, H. Tattegrain, C. Gabaude, Respiration and heart rate modulation due to competing cognitive tasks while driving, Frontiers in Human Neuroscience 12 (525) (2019) 1–8. doi:10.3389/fnhum.2018.00525.
[40] E. Pakdamanian, S. Sheng, J. A. Coan, L. Feng, Toward minimum startle after take-over request: A preliminary study of physiological data, in: 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '20 Adjunct), ACM, New York, NY, USA, 2020, pp. 27–30. doi:10.1145/3409251.3411715.
[41] T. Heine, G. Lenis, P. Reichensperger, T. Beran, O. Doessel, B. Deml, Electrocardiographic features for the measurement of drivers' mental workload, Applied Ergonomics 61 (2017) 31–43. doi:10.1016/j.apergo.2016.12.015.
[42] M. Shakouri, L. H. Ikuma, F. Aghazadeh, I. Nahmens, Analysis of the sensitivity of heart rate variability and subjective workload measures in a driving simulator: The case of highway work zones, International Journal of Industrial Ergonomics 66 (2018) 136–145. doi:10.1016/j.ergon.2018.02.015.
[43] D. D. Heikoop, J. C. de Winter, B. van Arem, N. A. Stanton, Effects of mental demands on situation awareness during platooning: A driving simulator study, Transportation Research Part F: Traffic Psychology and Behaviour 58 (2018) 193–209. doi:10.1016/j.trf.2018.04.015.
[44] B. Reimer, L. A. D'Ambrosio, J. F. Coughlin, Secondary analysis of time of day on simulated driving performance, Journal of Safety Research 38 (5) (2007) 563–570. doi:10.1016/j.jsr.2007.07.002.
[45] F. Naujoks, C. Mai, A. Neukum, The effect of urgency of take-over requests during highly automated driving under distraction conditions, in: T. Ahram, W. Karwowski, T. Marek (Eds.), Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE, Vol. 7, Kruków, Poland, 2014, pp. 431–438.
[46] S. M. Petermeijer, J. C. De Winter, K. J. Bengler, Vibrotactile displays: A survey with a view on highly automated driving, IEEE Transactions on Intelligent Transportation Systems 17 (4) (2016) 897–907. doi:10.1109/TITS.2015.2494873.
[47] M. R. Endsley, Situation awareness in future autonomous vehicles: Beware of the unexpected, in: S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, Y. Fujita (Eds.), Proceedings of the 20th Congress of the International Ergonomics Association, Vol. 824, Springer, Cham, 2019, pp. 303–309. doi:10.1007/978-3-319-96071-5{_}32.

B | LETTER OF APPROVAL ETHICS COMMITTEE

Date 11-02-2020
Contact person Ir. J.B.J. Groot Kormelink, secretary HREC
Telephone +31 152783260
E-mail j.b.j.grootkormelink@tudelft.nl

Human Research Ethics Committee
TU Delft
(http://hrec.tudelft.nl/)

Visiting address
Jaffalaan 5 (building 31)
2628 BX Delft

Postal address
P.O. Box 5015 2600 GA Delft
The Netherlands

Ethics Approval Application: Big Five personality traits and driving with automated driving systems - a simulator study
Applicant: Heikoop, Daniël

Dear Daniël Heikoop,

It is a pleasure to inform you that your application mentioned above has been approved.

Dear Daniël,

The information about the project is very thoroughly presented to the participants. HREC would like to suggest to you change the order of consent form and experiment information: i.e. put the consent-form at the start of the document.

Kind Regards,

Jan Salden

Acting Secretary HREC

Good luck with your research!

Sincerely,

Dr. Ir. U. Pesch
Chair HREC
Faculty of Technology, Policy and Management

C APPLICATION FORM

The following questions were asked in the application form which was made available through Qualtrix.

Table C.1: Application form personality experiment

	English	Dutch
	General	
1.	Do you own a driving license?	Bent u in het bezit van een rijbewijs?
2.	Will you be available from March to April to take part in our driving simulator experiment?	Bent u in maart tot april beschikbaar om mee te doen aan het rijsimulator experiment?
3.	I understand my participation is voluntary	Ik begrijp dat mijn deelname vrijwillig is
4.	I understand I am not automatically selected after filling in this questionnaire	Ik begrijp dat ik niet automatisch geselecteerd ben na het invullen van deze vragenlijst
	Contact details	
5.	Name and Surname	Voor- en achternaam
6.	E-mail address	E-mail adres
7.	Telephone number	Telefoonnumber
	Personal information	
8.	What is your gender?	Wat is uw geslacht?
9.	How old are you?	Wat is uw leeftijd?
10.	Which country are you from?	Uit welk land komt u oorspronkelijk?
11.	Which city do you currently live in?	In welke stad woont u?
12.	What is your profession?	Wat is uw beroep?
13.	What is your highest level of education?	Wat is uw hoogst afgeronde opleiding?
	Driving experience	
14.	For how many years have you been in possession of your driving license?	Hoe lang bent u al in het bezit van uw rijbewijs?
15.	How often did you drive in the last 12 months?	Hoe vaak reedt u in de afgelopen 12 maanden?
16.	How many kilometres did you drive in the last 12 months?	Hoeveel kilometer reedt u in de afgelopen 12 maanden?
17.	Do you have any experience with ADAS?	Heeft u enige ervaring met ADAS?
18.	Which ADAS did you use?	Welke ADAS heeft u wel eens gebruikt?
	Health questions	
19.	Do you use glasses or contact lenses?	Gebruikt u een bril of contactlenzen?
20.	Do you take any kind of drugs (including alcohol) at present?	Gebruikt u wel eens drugs (o.a. alcohol)?
21.	What kind of drugs do you normally take and how often do you take them?	Welke drugs gebruikt u en hoe vaak?
	Big Five Inventory (BFI)	
22.	Appendix D	

D BIG FIVE INVENTORY

The BFI listed in Table D.1 was used for selecting participants. Both an English and Dutch version was used. The inventory started with the following introduction:

English:
A number of characteristics that may or may not apply to you are described below. Please select for each statement a number indicating the extent to which you agree or disagree with that statement.

Dutch:
Hieronder staan een aantal persoonlijkheidskenmerken beschreven die mogelijk op u van toepassing zijn. Geef voor iedere stelling aan in hoeverre u het eens of oneens bent met die stelling.

A five-point Likert scale ranging from strongly disagree to strongly agree was used.

Table D.1: The 44-item BFI

	English	Dutch
1.	Is talkative	Spraakzaam is
2.	Tends to find fault with others	Geneigd is kritiek te hebben op anderen
3.	Does a thorough job	Grondig te werk gaat
4.	Is depressed, blue	Somber is
5.	Is original, comes up with new ideas	Origineel is, met nieuwe ideeën komt
6.	Is reserved	Terughoudend is
7.	Is helpful and unselfish with others	Behulpzaam en onzelfzuchtig ten opzichte van anderen is
8.	Can be somewhat careless	Een beetje nonchalant kan zijn
9.	Is relaxed, handles stress well	Ontspannen is, goed met stress kan omgaan
10.	Is curious about many different things	Benieuwd is naar veel verschillende dingen
11.	Is full of energy	Vol energie is
12.	Starts quarrels with others	Snel ruzie maakt
13.	Is a reliable worker	Een werker is waar men van op aan kan
14.	Can be tense	Gespannen kan zijn
15.	Is ingenious, a deep thinker	Scherpzinnig, een denker is
16.	Generates a lot of enthusiasm	Veel enthousiasme opwekt
17.	Has a forgiving nature	Vergevingsgezind is
18.	Tends to be disorganized	Doorgaans geneigd is tot slordigheid
19.	Worries a lot	Zich veel zorgen maakt
20.	Has a active imagination	Een levendige fantasie heeft
21.	Tends to be quiet	Doorgaans stil is
22.	Is generally trusting	Mensen over het algemeen vertrouwt
23.	Tends to be lazy	Geneigd is lui te zijn
24.	Is emotionally stable, not easily upset	Emotioneel stabiel is, niet gemakkelijk overstuur raakt
25.	Is inventive	Vindingrijk is
26.	Has an assertive personality	Voor zichzelf opkomt
27.	Can be cold and aloof	Koud en afstandelijk kan zijn
28.	Perseveres until the task is finished	Volhoudt tot de taak af is
29.	Can be moody	Humeurig kan zijn
30.	Values artistic, aesthetic experiences	Waarde hecht aan kunstzinnige en esthetische ervaringen
31.	Is sometimes shy, inhibited	Soms verlegen of geremd is
32.	Is considerate and kind to almost everyone	Attent en aardig is voor bijna iedereen
33.	Does things efficiently	Dingen efficiënt doet
34.	Remains calm in tense situations	Kalm blijft in gespannen situaties
35.	Prefers work that is routine	Een voorkeur heeft voor werk dat routine is
36.	Is outgoing, sociable	Hartelijk, een gezelschapsmens is
37.	Is sometimes rude to others	Soms grof tegen anderen is
38.	Makes plans and follows through with them	Plannen maakt en deze doorzet
39.	Gets nervous easily	Gemakkelijk zenuwachtig wordt
40.	Likes to reflect, play with ideas	Graag nadenkt, met ideeën speelt
41.	Has few artistic interests	Weinig interesse voor kunst heeft
42.	Likes to cooperate with others	Graag samenwerkt met anderen
43.	Is easily distracted	Gemakkelijk afgeleid is
44.	Is sophisticated in art, music, or literature	Het fijne weet van kunst, muziek of literatuur

E | ENGLISH - INFORMATION SHEET AND INFORMED CONSENT FORM

The informed consent form is sent to the participants after they are selected to participate in the driving simulation experiment. The form consists of two parts. First an information sheet with a detailed explanation of the experiment and procedure, the risks involved in participating, the privacy of the participants, sharing of results, ethical approval. The second part is the actual consent form for the previous mentioned details. The informed consent form is written in both English and Dutch. The Dutch version can be found in Appendix F.

TU Delft

Information sheet regarding the experiment and study + Informed consent form
February, 2020

1. Research group

1.1. Researchers in charge of the project

T. Ebbers	MSc. Student	Delft University of Technology
T. Marfoglia	MSc. Student	Delft University of Technology
M.P. Hagenzieker	Professor	Delft University of Technology
J.C.F. de Winter	Associate professor	Delft University of Technology
J.A. Annema	Assistant professor	Delft University of Technology
D.D. Heikoop	Post-doctoral researcher	Delft University of Technology

1.2. Organizations

Faculty of Civil Engineering and Geosciences, Department of Transport, Delft University of Technology.

This study is part of the research project 'Meaningful Human Control over Automated Driving Systems' (MHC-ADS) of the Department of Transport and Planning, Delft University of Technology.

2. This document

This informed consent document consists of two parts:
1) Information sheet
2) Informed consent form

You are asked to read this document carefully before signing the informed consent form. Information is provided regarding the purpose of this study, your participation, the procedure of the experiment, the expected benefits, risks associated with this experiment, information regarding data protection, privacy and confidentiality, the sharing of the results, and who are responsible for this study. If something is unclear or needs additional explanation, please contact any of the researchers. After reading the information sheet and if all questions or concerns are answered, you can choose to participate in this study. To participate, please fill in the informed consent form on the last page of this document. Your signature is required for participation.

TU Delft

3. Purpose of this research

An increasing amount of vehicles are already equipped with longitudinal and lateral automated support up to SAE level 2 (SAE, 2018). Nowadays, SAE level 3 automated vehicles are the next step in which the human driver is still a fallback-ready user in case the vehicle is requesting it. This part, also called the transition of control, whereby the vehicle needs to interact with the human driver to take over control is interesting to research, especially when human drivers differ in personality traits. A driving simulator study with around 100 participants will be performed regarding the time it takes a driver to get full control of the vehicle. Moreover, this will be done by speech-based auditory feedback in several stages of urgency. Various measurements will be carried out for the aim of this research and will be analyzed and published in order to contribute to the research in the interaction between the vehicle and the human driver.

4. Participation

4.1. Location of the experiment

The experiment will be held at the faculty of Civil Engineering and Geosciences at Delft University of Technology: Stevinweg 1, 2628CN, Delft. The driving simulators are located on the 4th floor, room 4.32.6.

4.2. Eligibility criteria

You are invited to participate in this experiment if:

- You are 18 years or older
- You have a car driving license
- You are not under the influence of drugs, alcohol or other substances that compromise your driving ability.
- You have not experienced (severe) simulation or motion sickness.

The researchers reserve the right at any time to refuse or excuse (from an in-progress session) any participant who meets/no longer meets the study requirements or who is behaving in an unnecessarily unsafe manner.

4.3. Voluntary participation and the right to refuse or withdraw

Participating in this study is completely voluntary. If you have any questions or concerns regarding this study, please contact one of the researchers. If you do agree to participate in this study, you can withdraw at any moment without comment or penalty. Withdrawal from the experiment is possible until 10 working days after completing the experiment. In case of withdrawal, all personal data will be removed from this study.
Participants will be given the opportunity to get insight into their own data obtained in this experiment - ask any of the researchers to provide you with this data.
Rectification of the data is not possible.

5. Procedure

This study consists of one driving simulator experiment. The experiment focuses on the personality of drivers and their behavior in automated vehicles. In the experiment, transitions of control between the automated driving system and the driver will be simulated. Data regarding how drivers experience these control transitions, and their according to driving behavior, will be collected. The data will be collected by the driving simulator, a camera and by sensors mounted on the fingers of the participants to measure heart rate and electrical conductivity of the skin.

5.1. Experiment

You will be asked to perform one driving sessions of approx. 30 minutes in a highway setting. Data from this experiment will be used to analyze the effect of personality on the experience of transitions of control in automated vehicles. The simulated vehicle is a generic sedan car. The simulated vehicle is controlled in the same way as a normal car with automatic gearbox: it has pedals and turn signals. Furthermore, the dashboard of the vehicle will be simulated showing the turn signals, speedometer and tachometer. Also side view mirrors and a rearview mirror are simulated.
The following data will be collected: steering and pedal input, eye movement, heart rate, skin conductance level.

5.2. Prior to the simulator sessions

Prior to the simulator sessions, this information sheet with the informed consent form will be sent to you. Furthermore, you are asked to fill in a demographic questionnaire, a driving experience questionnaire, a healthiness questionnaire, and the Big Five personality test.
Once at the experiment location, a safety instruction will be given on operating the driving simulator.

5.3. Practice simulator session

The experiment includes a practice round, in which you can get familiar with the driving simulator like the virtual environment and the steering wheel and pedals. This practice round will take around 5 minutes in which you have some freedom to drive around and to follow some instructions.

5.4. Simulator session instructions

5.4.1. Driving

In the experiment, you will be asked to drive as normally as you are allowed to do in normal driving conditions with respect to traffic regulations. Moreover, you will be driving in the utmost right lane on a three-lane highway and are allowed to take-over slower vehicles if circumstances permit this.

TU Delft

5.4.2. Controls

In the first part of the scenario, the vehicle is driving autonomously on the highway at 100 km/h. During this part of the scenario, no input of the human driver is necessary (like steering or gas pedal input). The vehicle could ask at a certain moment to take-over control of the vehicle, in which you have to put your hands on the steering wheel and feet on the pedals.

5.4.3. Scenario

In the scenario, other vehicles will be driving around on the highway and you need to treat them just like you would do in a normal driving situation. Moreover, the scenario allows overtaking other vehicles and driving faster than the maximum speed. You are asked to drive like the normal driving rules you know, so use your direction indicator and do not drive faster than allowed.
The full scenario is a long stretch of a highway, in which several turns are included. You are driving first in automated mode, while at a certain time the vehicle is requesting to take-over control and drive further in manual mode till the vehicle again informs that it will take over the driving task from you.

5.5. Duration and time commitment

The experiment will take around 60 minutes, which includes the welcome, signing the consent form, getting familiar with the driving simulator in the test round and filling in the questionnaire.

6. Expected benefits

The outcome of this experiment will be used for the research into automated vehicles. It will not directly benefit you as a driver immediately, but it will improve the understanding of automation and the interaction between a vehicle and a human driver. Your contribution to this project will help make automated vehicles in the future even more likely and better.

7. Risks associated with participation

In the simulator, participants may experience simulator motion sickness. The experiment can be stopped immediately if necessary, by the participant or by the researcher. Furthermore, the participant needs to wear the seatbelt during the experiment. Taking off the seatbelt during the experiment will cause the test to stop.
If the participant loses control over the vehicle, this can result in an accident in the virtual scenario. This does not harm the participant physically, but it can be emotionally demanding. To overcome these types of problems, no other persons are visible in the scenario and other vehicles are non-solid objects (the participant can drive through them).
The simulator is located in a small room at the faculty CiTG of the TU Delft, in which no mechanical ventilation is available. To get more airflow in the room, a fan will be used. Participants can fall over the cables of this fan or from the simulator itself. But due to the fact, these cables are stuck on the floor by tape, the chance is small. During the experiment,

TU Delft

temporary cables can be necessary to fulfill the simulator study, so before the participant can leave the simulator, the researcher needs to take these cables away.

8. Privacy and confidentiality

All data collected in this study will be stored securely as of the Data Management policy of Delft University of Technology. Only the researchers involved in this study can access the data. Data will be stored, encrypted and pseudonymised, on the TU Delft server. This data will be stored for 10 years. The non-identifiable data from this study will be stored in an open-access database for future use in comparative studies and for secondary analysis.

9. Sharing of results

The results of the study are presented in the research reports of the researcher. Moreover, the study might be presented in a scientific journal. The data of the driving simulator could be used in follow-up research into this field like in related studies, simulator training and the design of vehicles.

10. Responsibility

The researchers and the institution involved in this research are not responsible for any damages during the travel to or from the location of the experiment.

11. Questions/further information about the project

If you have any questions or comments regarding the study and the experiment, or if you require further information, please contact one of the researchers:

Researchers	*E-mail addresses*	*Telephone numbers*
T. Ebbers	t.ebbers@student.tudelft.nl	+31 (0)630989569
T. Marfoglia	t.marfoglia@student.tudelft.nl	+31 (0)681220906
M.P. Hagenzieker	m.p.hagenzieker@tudelft.nl	
J.C.F. de Winter	j.c.f.dewinter@tudelft.nl	
J.A. Annema	j.a.annema@tudelft.nl	
D.D. Heikoop	d.d.heikoop@tudelft.nl	

12. Ethical approval and complaints regarding the conduct of the project

This study will be approved by the Human Research Ethics Committee (HREC) of the TU Delft. A verification of this approval can be obtained by sending an email to HREC@tudelft.nl. If you have any complaints or suggestions about the ethical conduct of this project, please contact HREC by sending an email to the above-mentioned email address.

TU Delft

Consent form

Please tick the appropriate boxes	Yes	No
Taking part in the study		
I have read and understood the study information dated Februari 2020, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	□	□
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	□	□
I understand that taking part in the study involves collecting data like video-recording, which will be transcribed as text, and completing questionnaires.	□	□
Risks associated with participating in the study		
I understand that taking part in the study involves the following risks: motion sickness due to the experiment in a driving simulator. Emotional discomfort when experiencing a virtual accident.	□	□
Use of the information in the study		
I understand that the information I provide will be used for Master's theses, conference presentations and articles in scientific journals.	□	□
I understand that personal information collected about me that can identify me, such as [e.g. my name or where I live], will not be shared beyond the study team.	□	□

TU Delft

Future use and reuse of the information by others

I give permission for the data obtained with the sensors, camera and driving simulator, as well as all data from the questionnaires that I provide to be archived in the TU Delft repository so it can be used for future research and learning. ☐ ☐

Signatures

_________________________ _______________________ _____________

Name of participant Signature Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands what they are freely consenting.

_________________________ _______________________ _____________

Researcher name Signature Date

_________________________ _______________________ _____________

Researchers name Signature Date

F DUTCH - INFORMATION SHEET AND INFORMED CONSENT FORM

Document met informatie over het experiment en het daarbij behorende toestemmingsformulier.
Februari, 2019

1. Onderzoeksgroep

1.1. Onderzoekers verantwoordelijk voor het project

T. Ebbers	MSc. Student	Delft University of Technology
T. Marfoglia	MSc. Student	Delft University of Technology
M.P. Hagenzieker	Professor	Delft University of Technology
J.C.F. de Winter	Associate professor	Delft University of Technology
J.A. Annema	Assistant professor	Delft University of Technology
D.D. Heikoop	Postdoctoraal onderzoeker	Delft University of Technology

1.2. Organisaties

Faculteit van Civiele Techniek en Geowetenschappen, afdeling Transport, Technische Universiteit Delft.

Deze studie is onderdeel van het onderzoek 'Meaningful Human Control over Automated Driving Systems' (MHC-ADS) van de afdeling Transport & Planning, Technische Universiteit Delft.

2. Inhoud van dit document

Dit toestemmingsformulier bestaat uit twee delen, namelijk:
1) Informatieformulier
2) Toestemmingsformulier

U wordt gevraagd om dit document zo zorgvuldig mogelijk door te lezen voordat u het toestemmingsformulier ondertekend. De verstrekte informatie bevat het doel van dit onderzoek, uw deelname, de procedure, de verwachte voordelen, de risico's, informatie met betrekking tot gegevensbescherming, privacy en vertrouwelijke informatie. Ook wordt u geïnformeerd over het delen van de informatie en wie verantwoordelijk zijn voor dit onderzoek. Als u na het lezen van dit informatiedocument nog vragen heeft of extra uitleg wilt hebben, kunt u contact opnemen met een van de onderzoekers. Als na het lezen van het informatiedocument al uw vragen zijn beantwoord, kunt u er voor kiezen om mee te doen aan het onderzoek. Om uw deelname te bevestigen, vult u het toestemmingsformulier op de laatste pagina van dit document in. Uw handtekening is vereist voor deelname.

3. Doel van het onderzoek

Tegenwoordig zijn een toenemend aantal voertuigen reeds uitgerust met geautomatiseerde rijhulpsystemen tot SAE level 2 (SAE, 2018). De volgende stap in geautomatiseerde voertuigen is SAE level 3, waarbij de menselijke bestuurder nog steeds verantwoordelijk is voor de rijtaak en gereed moet zijn om de rijtaak over te nemen in het geval het voertuig hierom vraagt. Dit wordt ook wel de overgang van controle genoemd, waarbij het voertuig vraagt aan de menselijke bestuurder om de controle over het voertuig over te nemen. Deze overgang van controle is interessant voor dit onderzoek, samen met het onderscheiden van menselijk bestuurders met verschillende persoonlijkheidskenmerken. Een rijsimulator onderzoek met ongeveer 100 deelnemers zal worden uitgevoerd, waarbij gekeken wordt hoe lang een menselijke bestuurder nodig heeft om volledige controle over het voertuig te krijgen na een overgang van controle. Deze overgang zal in gang worden gezet door een op spraak gebaseerde auditieve feedback in verschillende fasen van urgentie. Voor het onderzoek zullen verschillende metingen verricht worden waarna deze geanalyseerd en gepubliceerd worden om bij te dragen aan het onderzoek naar de interactie tussen voertuig en de menselijke bestuurder.

4. Participation

4.1. Locatie van het experiment

Het experiment zal plaatsvinden op de faculteit Civiele Techniek en Geowetenschappen op de Technische Universiteit Delft: Stevinweg 1, 2628CN, Delft. De rijsimulatoren bevinden zich op de vierde verdieping in ruimte 4.32.6.

4.2. Toelatingscriteria

U mag deelnemen aan dit experiment, als:
- U bent 18 jaar of ouder.
- U bent in het bezit van een rijbewijs.
- U bent niet onder invloed van drugs, alcohol, of andere stoffen die uw rijvaardigheid beïnvloeden.
- U heeft nog nooit (heftige) simulatie- of bewegingsziekte ervaren.

De onderzoekers behouden zich te allen tijde het recht voor om een deelnemer die niet (langer) voldoet aan de studie-eisen te weigeren (van een lopende sessie) of die zich op een onnodig onveilige manier gedraagt.

4.3. Vrijwillige deelname en recht tot terugtrekking

Deelname aan dit onderzoek is volledig vrijwillig. Als u vragen of opmerkingen heeft over dit onderzoek, kunt u contact opnemen met een van de onderzoekers. Als u akkoord gaan met deelname aan dit onderzoek, kunt u zich op elk gewenst moment terugtrekken zonder gevolgen en zonder verantwoording. Terugtrekken van dit onderzoek is mogelijk tot 10 werkdagen na deelname aan het experiment. In geval van terugtrekking van dit experiment, zullen alle persoonlijke data van het onderzoek verwijderd worden.
Aan deelnemers aan dit onderzoek wordt de mogelijkheid tot inzicht in eigen data van deze studie worden gegeven - vraag een van de onderzoekers om u deze informatie te verschaffen.

Het is niet mogelijk om uw data te veranderen.

5. Procedure

Een rijsimulator experiment is onderdeel van deze studie. Het experiment focust zich op de persoonlijkheid van bestuurders en hun gedrag in autonome voertuigen. Het experiment simuleert transities van controle tussen het autonome rijsysteem en de bestuurder. Data zal worden verzameld met betrekking tot hoe bestuurders deze transities van controle ervaren en hun overeenkomstige rijgedrag. De data zal worden verzameld met behulp van de rijsimulator, een camera, en sensoren op de vingers van de deelnemers om de hartslag en elektrische geleidbaarheid van de huid te meten.

5.1. Experiment

U wordt gevraagd om deel te nemen aan één rij sessie van ongeveer 30 minuten in een snelweg setting. Data van het experiment zal worden gebruikt om het effect van persoonlijkheid te meten op hoe transities van controle in autonome voertuigen worden ervaren. Het gesimuleerde voertuig betreft een standaard sedan voertuig. Het gesimuleerde voertuig wordt net zo bestuurd als een normaal voertuig met automaat: het is ook uitgerust met de standaard pedalen en richtingaanwijzers. Het dashboard van het gesimuleerde voertuig laat de richtingaanwijzers, snelheidsmeter en tachometer zien. Ook zij-spiegels en een achteruitkijkspiegel zijn gesimuleerd.
De volgende data zal worden verzameld: stuur- en pedaal-input, oogbewegingen, hartslag en elektrische geleidbaarheid van de huid.

5.2. Voorafgaand aan het simulator onderzoek

Voor het simulator onderzoek zal dit toestemmingsformulier met informatie over het onderzoek naar u worden gestuurd. U wordt verder nog gevraagd om een aantal vragenlijsten in te vullen, dit zijn: een demografische vragenlijst, een vragenlijst over uw rijervaring, een gezondheidsvragenlijst en ten slotte de Big Five persoonlijkheidstest.
Eenmaal op de experiment-locatie zal een veiligheidsinstructie worden gegeven betreffende het rijden in een rijsimulator.

5.3. Oefen-sessie rijsimulator

Het experiment omvat ook een oefensessie, waar u zich kunt familiariseren met de rijsimulator, dit betreft dan de virtuele omgeving, het stuur en de pedalen van de rijsimulator. Deze oefensessie zal ongeveer 5 minuten kosten. Tijdens deze sessie heeft u de vrijheid om vrij te rijden, maar zult u ook enkele instructies moeten opvolgen.

5.4. Instructies simulator sessie

5.4.1. Het rijden

Tijdens het experiment zal u worden gevraagd om zo normaal als mogelijk te rijden, zoals u wordt mogelijk gemaakt onder normale rij-omstandigheden en met inachtneming van de verkeersregels. Verder wordt u gevraagd om in de rechterbaan te rijden van de drie-baans snelweg. U wordt toegestaan om langzaam rijdend verkeer in te halen als de rij-omstandigheden dit mogelijk maken.

5.4.2. Bediening

Gedurende het eerste gedeelte van het scenario, zal het voertuig autonoom op de snelweg rijden met een snelheid van 100 km/u. Gedurende dit gedeelte van het scenario is geen enkele input nodig van de bestuurder (dus geen stuur of pedaal input). Het voertuig kan u op elk moment vragen om de controle over te nemen van het autonome rijsysteem, waarbij u uw handen op het stuur moet plaatsen en de voeten op de pedalen.

5.4.3. Scenario

In het scenario zullen ook andere voertuigen rijden op de snelweg, u wordt gevraagd hiermee om te gaan zoals u gewoonlijk doet tijdens het rijden. U wordt toegestaan om andere voertuigen in te halen en sneller te rijden dan de maximaal toegestane snelheid. Echter wordt u wel gevraagd zich te houden aan de verkeersregels, richtingaanwijzers te gebruiken en, indien mogelijk, zich te houden aan de maximaal toegestane snelheid.
Het scenario bestaat uit een lang stuk snelweg met een aantal bochten. Het voertuig zal beginnen in autonome modus. Na een bepaalde tijd wordt u gevraagd om controle over te nemen van het voertuig, waarna u in handmatige modus zal rijden tot het voertuig u informeert dat het autonome systeem het rijden van u zal overnemen.

5.5. Duur van het onderzoek

Het onderzoek zal in totaal ongeveer 60 minuten in beslag nemen, dit omvat het invullen van de vragenlijsten en het experiment: onthaal, ondertekenen van het toestemmingsformulier, de oefensessie en het daadwerkelijke experiment.

6. Verwachte resultaten

De resultaten van dit experiment zullen worden gebruikt voor het onderzoek naar geautomatiseerde voertuigen. Deze resultaten zullen u als bestuurder niet direct ten goede komen, maar zal het bijdragen aan het verbeteren van de interactie tussen een voertuig en de menselijke bestuurder. Uw deelname aan dit project zal er aan bijdragen dat geautomatiseerde voertuigen beter worden en de introductie ervan op de openbare weg in de toekomst nog waarschijnlijker zullen zijn.

7. Risico's verbonden aan deelname

Deelnemers kunnen in de simulator bewegingsziekte ervaren. Het experiment kan indien nodig onmiddellijk worden beëindigd door de deelnemer of de onderzoeker. Daarnaast moet de deelnemer te allen tijde de veiligheidsgordel dragen. Als u tijdens het experiment de veiligheidsgordel ontgrendeld, zal de test worden gestopt. De deelnemer kan tijdens het experiment de macht over het stuur verliezen, wat kan leiden tot een virtueel ongeval. Dit is fysiek niet schadelijk voor bestuurder, maar kan emotioneel zwaar zijn. Om deze problemen te voorkomen zijn er geen andere personen zichtbaar in het scenario en zijn andere voertuigen niet solide objecten (de deelnemer kan door andere voertuigen heen rijden).
De rijsimulator bevindt zich in een kleine ruimte in de faculteit CiTG van de TU Delft. In deze ruimte is geen mechanische ventilatie aanwezig, maar door het gebruik van een ventilator zal er een voldoende luchtstroom aanwezig zijn. Deelnemers kunnen over kabels van de ventilator of van de rijsimulator vallen. Doordat deze kabels op de vloer zijn vastgetaped is

deze kans zeer klein. Tijdens het experiment kunnen tijdelijke kabels worden gebruikt maar zullen worden verwijderd door de onderzoeker voordat de deelnemer de rijsimulator verlaat.

8. Privacy en vertrouwelijkheid

Alle verzamelde data van deze studie zal beveiligd worden opgeslagen volgens het Data Management beleid van de Technische Universiteit Delft. Alleen de onderzoekers die betrokken zijn bij dit onderzoek hebben toegang tot de data. De data zal worden worden gecodeerd en gepseudonimiseerd en zal worden bewaard op de server van de TU Delft. Deze data zal vervolgens worden opgeslagen voor de komende 10 jaar. Alle niet-identificeerbare data van deze studie zal worden opgeslagen in een vrij-toegankelijke database voor toekomstig gebruik in vergelijkbare studies of voor toekomstige analyse.

9. Delen van resultaten

De resultaten van het onderzoek worden gepresenteerd in de onderzoeksrapporten van de onderzoekers. Daarnaast kan het onderzoek worden gepubliceerd in een wetenschappelijk tijdschrift. De gegevens van de rijsimulator kunnen worden gebruikt in een vervolgonderzoek op dit gebied zoals gerelateerde studies, simulator trainingen of voor het ontwerp van geautomatiseerde voertuigen.

10. Verantwoordelijkheid

De onderzoekers die bij dit onderzoek betrokken zijn zijn niet verantwoordelijk voor eventuele schade tijdens de reis naar of van de locatie van het onderzoek.

11. Vragen of verdere informatie over dit onderzoek

Als u nog vragen of opmerkingen heeft over dit onderzoek, of als u nog andere informatie nodig heeft, kunt u contact opnemen met een van de volgende onderzoekers:

Onderzoekers	*E-mail adressen*	*Telefoonnummers*
T. Ebbers	t.ebbers@student.tudelft.nl	+31 (0)630989569
T. Marfoglia	t.marfoglia@student.tudelft.nl	+31 (0)681220906
M.P. Hagenzieker	m.p.hagenzieker@tudelft.nl	
J.C.F. de Winter	j.c.f.dewinter@tudelft.nl	
J.A. Annema	j.a.annema@tudelft.nl	
D.D. Heikoop	d.d.heikoop@tudelft.nl	

12. Ethische goedkeuring en klachten met betrekking tot de uitvoering van het project

Het onderzoek is goedgekeurd door de Human Research Ethics Committee (HREC) van de TU Delft. Een verificatie van deze goedkeuring kan worden aangevraagd door een e-mail te sturen naar HREC@tudelft.nl. Als u klachten of suggesties heeft over het ethische gedrag van dit onderzoek, neem dan contact op met HREC door een e-mail te sturen naar het bovengenoemde e-mailadres.

Toestemmingsformulier

Vink de vakjes die op u van toepassing zijn aan	Ja	Nee
Deelname aan het onderzoek		
Ik heb alle informatie betreffende het onderzoek d.d. februari 2020 gelezen en begrepen, of iemand heeft dit mij voorgelezen. Ik heb de mogelijkheid gekregen om vragen te stellen over het onderzoek en mijn vragen zijn naar tevredenheid beantwoord.	□	□
Ik stem ermee in vrijwillig deel te nemen aan dit onderzoek en begrijp dat ik kan weigeren om vragen te beantwoorden en ik kan me op elk moment terugtrekken uit het onderzoek, zonder een reden op te geven.	□	□
Ik begrijp dat deelname aan het onderzoek het verzamelen van gegevens omvat, zoals video-opnames, die als tekst wordt getranscribeerd, en het invullen van vragenlijsten.	□	□
Risico's verbonden aan deelname aan dit onderzoek		
Ik begrijp dat deelname aan het onderzoek de volgende risico's met zich meebrengt: bewegingsziekte door het experiment in een rijsimulator, emotioneel ongemak bij het ervaren van een virtueel ongeval.	□	□
Gebruik van informatie uit dit onderzoek		
Ik begrijp dat de informatie die ik geef zal worden gebruikt voor masterscripties, congres presentaties en artikelen in wetenschappelijke tijdschriften.	□	□
Ik begrijp dat persoonlijke informatie die over mij is verzameld die mij kunnen identificeren, zoals [bijv. mijn naam of waar ik woon], niet wordt gedeeld buiten het onderzoeksteam.	□	□

Toekomst (her)gebruik van informatie door derden

Ik geef toestemming om de gegevens verkregen door middel van de sensoren, camera en rijsimulator, evenals alle gegevens uit de vragenlijsten, te archiveren in de TU Delft data-opslag zodat toekomstige onderzoeken en studies deze data kunnen gebruiken.	□	□

Handtekeningen

________________________ ______________________ _____________

Naam van deelnemer Handtekening Datum

Ik heb, indien nodig, het informatie document zorgvuldig voorgelezen aan de potentiële deelnemer en verklaar hierbij, voor zover ik kan, dat ik gezorgd heb dat de deelnemer begrijpt waarmee hij/zij instemt.

________________________ ______________________ _____________

Naam onderzoeker Handtekening Datum

________________________ ______________________ _____________

Naam onderzoeker Handtekening Datum