TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Ergonomie

Take-over Performance in Conditionally Automated Driving: Effects of the Driver State and the Human-Machine-Interface

Dipl.-Ing. (Univ.) Jonas Radlmayr

Vollständiger Abdruck der von der Fakultät für Maschinenwesen der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:	Prof. Dr.-Ing. Boris Lohmann
Prüfer der Dissertation:	1. Prof. Dr. phil. Klaus Bengler
	2. Prof. Dr. phil. Mark Vollrath

Die Dissertation wurde am 08.06.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Maschinenwesen am 21.10.2020 angenommen.

Acknowledgement/Danksagung

Die Dissertation entstand während meiner Zeit am Lehrstuhl für Ergonomie der TUM, für die ich Herrn Prof. Bengler außerordentlich dankbar bin. Das hohe Vertrauen, die Freiheit, Themen und Projekte eigenverantwortlich zu bearbeiten, die Arbeitskultur am Lehrstuhl und die Diskussionen haben mich stark geprägt und waren entscheidende Grundlage für den erfolgreichen Abschluss der Promotion. Bei Prof. Lohmann und Prof. Vollrath bedanke ich mich für die Übernahme von Vorsitz und Zweitkorrektur und das zügige und problemlose Verfahren.
Die Dissertation ist zu einem großen Teil geprägt vom Austausch im Projekt Ko-HAF. Für die konstruktive Zusammenarbeit und die Unterstützung bin ich allen Mitgliedern des Projekts sehr dankbar, vor allem Thorsten Stürmer und Sebastian Krug.
Während der Zeit am Lehrstuhl hatte ich das große Glück, viele hilfsbereite Menschen kennenzulernen. Bei meinen Studenten bedanke ich mich für ihre Mühen und Arbeit in diversen Studienarbeiten, vor allem bei Lisa Scherer und Fabian Fischer. Bei Lutz Lorenz bedanke ich mich sehr herzlich für seine Bereitschaft, mein Mentor zu sein. Ein großer Dank geht an beide Sekretariate und die Werkstatt für ihre Hilfsbereitschaft in allen Belangen. Neben der Unterstützung, die ich in Projekten und der Promotion erfahren habe, haben mich vor allem der Austausch mit großartigen Kolleginnen und Kollegen jeden Tag gerne an den Lehrstuhl kommen lassen. Ein großes Dankeschön an Andreas Haslbeck, Christoph Rommerskirchen, Markus Zimmermann, Bastiaan Petermeijer, Phillip Kerschbaum, Veronika Wehlack, Jonas Schmidtler, Patrick Gontar, Martin Götze, Ingrid Bubb, Lorenz Prasch, Philipp Maruhn, Alexander Feierle, Michael Rettenmaier, Andrea Müller und Deike Albers. Ein spezieller Dank geht an Moritz Körber für das sorgfältige Korrekturlesen, dem Bekanntmachen von wirklich hilfreichen R-Paketen und die beständige Unterstützung in statistischen Fragen.
Meinen Bürokollegen Anna Feldhütter und Yucheng Yang danke ich für eine einmalige Zeit, die tiefen und weniger tiefen Gespräche, die gemeinsamen Konferenzreisen, die jederzeit offenen Ohren für Probleme und Sorgen und ihre Freundschaft über den Lehrstuhl hinaus.
Hervorheben möchte ich Christian Gold, der mich in meiner Diplomarbeit mit dem Thema bekannt gemacht hat, mir den Start am Lehrstuhl und in Ko-HAF ermöglicht hat, ein großartiger Bürokollege und Korrekturleser war und der in vielen Belangen ein großes Vorbild ist.
Mein Dank gebührt auch meinen Freunden Christoph Jenke und Nicolas Brossardt für die Motivation weiter zu schreiben und zu trainieren und dabei nicht die wirklich wichtigen Dinge des Lebens außer Acht zu lassen.
Besonders meiner Familie bin ich zu Dank verpflichtet, die mich immer unterstützt hat und stete Quelle an Ansporn und Motivation war, die Dissertation abzuschließen.
Der weitaus größte Dank gilt meiner Frau Marlit. Ohne dein Verständnis und deine Geduld, die jahrelange Unterstützung in allen Situationen, die bedingungslose Hilfe und Motivation vor allem in den schwierigsten Zeiten wäre diese Dissertation sicher nicht entstanden. Vielen Dank!

Jonas Radlmayr

Abstract

While partial driving automation is already available, conditionally automated driving (CAD) in commercial vehicles is imminent. The paradigm change of CAD incorporates drivers leaving the driver-vehicle control loop, but remain fallback-ready users in case of system limits. This take-over and associated human factors challenges are of crucial interest for the safety and comfort of CAD and have seen great interest in science and industry. Based on an extensive literature research on past findings and relevant effects, the research questions of this thesis were derived. They address the effect of the driver state and the human-machine interface (HMI) on the take-over performance in CAD.

The empirical basis of this thesis comprises four experiments. The first experiment offers comprehensive insight about driver state changes caused by prolonged periods of automated driving. Results showed a strong influence of situational factors, but no influence of driver state changes on take-over performance. Experiment 2 focused on state changes caused by the engagement in different non-driving related tasks (NDRTs). Results allowed a detection of significant driver state changes which did not influence the take-over performance. Situational effects on take-over performance were revealed again, confirming the findings from the first experiment.

To allow a deeper understanding of relevant effects in combination with individual differences of drivers on take-over performance, the data from both experiments were pooled together in a modeling approach. Mixed models were chosen to quantify idiosyncratic effects in addition to improve prediction quality of take-over performance with regard to known effects. Linear mixed models were fitted for the take-over time and the time to collision, while bi- and multinomial mixed logistic regression was applied for crash probability and longitudinal and lateral accelerations. Results showed strong situational effects on take-over performance while state changes, e.g. visual attention, showed limited effects on driver behavior. Take-over time and braking behavior was highly affected by driver predispositions while the time to collision and lateral accelerations were hardly influenced by individual differences in the models.

Based on the findings from Experiments 1 and 2 and the modeling approach, the HMI was optimized. The third experiment evaluated the possibility of peripheral monitoring during CAD to improve visual perception and the situation awareness of drivers. Results showed no significant effects on take-over performance. Findings, that visual NDRTs during CAD lead to a significant reduction of visual perception and situation awareness were supported, but the possibility of peripheral monitoring presented no feasible solution. The fourth experiment provided insights into offering additional visual information on the reason for the take-over and the system limit during the take-over process. Take-over performance was unaffected by the optimized HMI whereas the subjective ratings of the take-over and the HMI benefited significantly. Participants reported significantly higher values for perceived safety, usability, intention to use and satisfaction compared to a generic HMI.

Concluding, this thesis provides an empirical comparison of various effects on take-over performance, including situational parameters, driver state and the HMI. While these results underline current findings in the literature, the quantification of idiosyncratic effects utilizing mixed models offers a novel and more comprehensive understanding of human factors challenges concerning the take-over in CAD.

Zusammenfassung

Hochautomatisiertes Fahren (HAF), in dem eine Automation die Fahrzeugführung übernimmt, sodass sich Fahrer mit anderen, fahrfremden Tätigkeiten (ffT) beschäftigen dürfen, soll demnächst verfügbar sein. Die Automation kann allerdings Systemgrenzen erreichen, in denen die Fahrer als Rückfallebene wieder manuell fahren müssen, sie "übernehmen". Der Paradigmenwechsel, dass Fahrer während aktiver Hoch-Automation im Vergleich zum teilautomatisierten Fahren nicht mehr überwachen müssen, in Kombination mit der Übernahme, beinhaltet eine Reihe neuer, ergonomischer Fragestellungen. Die Forschungsfragen dieser Arbeit adressieren den Einfluss der Verkehrssituation, des Fahrerzustands während HAF, des Anzeigekonzepts und den Einfluss von individuellen Unterschieden auf die Übernahmeleistung an Systemgrenzen.

Zentraler Kern der Arbeit sind vier Fahrsimulationsversuche, in denen Daten zur Beantwortung der Forschungsfragen erhoben wurden. Der erste Versuch untersuchte den Einfluss längerer Automationsdauern und zeigte keinen Einfluss des Fahrerzustands, aber einen starken Effekt der Situation auf die Übernahmeleistung. Im zweiten Versuch wurden unterschiedliche ffT untersucht, die ebenfalls keinen Einfluss zeigten, wohingegen Situationseinflüsse abermals signifikant die Übernahmeleistung beeinflusst haben.

Aufgrund des ähnlichen Versuchsdesigns der ersten beiden Versuche wurden die Daten für eine kombinierte Modellierung herangezogen. Übernahmezeit, Zeit bis zu einer potenziellen Kollision (time to collision, TTC) und auftretende Beschleunigungen wurden mithilfe von gemischten Modellen (mixed models) modelliert, die gleichzeitig zu den Einflüssen von Situation, Fahrerzustand, ffT, etc. die Quantifizierung von individuellen Unterschieden zwischen Fahrern erlaubten. Der starke Situationseinfluss aus den Versuchen 1 und 2 wurde bestätigt, zudem hat die Verkehrsdichte, das Alter und die visuelle Aufmerksamkeit einen Einfluss auf ausgewählte Metriken. Die Quantifizierung der individuellen Einflüsse zeigte einen starken Effekt dahingehend, dass Fahrer individuell unterschiedlich schnell übernahmen und ein stark individuelles Bremsverhalten zeigten, wohingegen die laterale Beschleunigung und die TTC kaum individuelle Unterschiede zeigten.

Auf Basis der Modellierungs- und Versuchsergebnisse wurde in Versuch 3 das Anzeigekonzept optimiert, um Fahrern während aktiver Automation die Möglichkeit der peripheren Überwachung zu geben. Die Ergebnisse zeigten keine signifikante Verbesserung der Übernahmeleistung, wobei Fahrer ohne visuelle ffT das größte Situationsbewusstsein vor der Übernahme aufwiesen. Im vierten Versuch wurden die Fahrer während der Übernahme durch die Anzeige der Systemgrenze im Head-Up Display und dem Grund für die Übernahme unterstützt. Auch hier zeigte sich kein Einfluss auf die Übernahmeleistung, allerdings schnitt die subjektive Bewertung der Übernahme und des neuartigen Anzeigekonzepts deutlich besser ab als das generische Anzeigekonzept aus den Versuchen 1-3. Vor allem das Sicherheitsempfinden, die Zufriedenheit und die Gebrauchstauglichkeit konnten signifikant verbessert werden.

Die Arbeit liefert einen wichtigen Beitrag zur Bewertung von Einflüssen wie Situation, Fahrerzustand und des Anzeigekonzepts auf die Übernahme beim HAF und erweitert damit die bestehende Forschungslandschaft. Die Quantifizierung der individuellen Unterschiede zwischen Fahrern und deren Vergleich mit den bestehenden Einflüssen leistet einen entscheidenden, neuen Beitrag zu einem umfassenden Verständnis der ergonomischen Herausforderungen der Übernahme beim HAF.

Glossary

AIC	Akaike Information Criteria
ANOVA	Analysis of Variance
AOI	Area of Interest
BASt	Bundesanstalt für Straßenwesen German Federal Highway Research Institute
CAD	Conditionally Automated Driving
CID	Central Information Display
COP	Center of Pressure
DDT	Dynamic Driving Task
GG	Greenhouse-Geisser-correction
GLM	Generalized Linear Model
GLMM	Generalized Linear Mixed Model
HGD	Horizontal Gaze Dispersion
HMI	Human-Machine Interface
HUD	Head-Up Display
ICC	Intraclass Correlation Coefficient
KSS	Karolinska Sleepiness Scale
LRT	Likelihood Ratio Test
M	Mean
MMLR	Multinomial Mixed Logistic Regression
MWWT	Mann-Whitney-Wilcoxon Test
NDRA/T	Non-driving related activities or tasks
ODD	Operational Design Domain
OEDR	Object and Event Detection and Response
OLS	Ordinary Least Squared (regression)
PAD	Partially Automated Driving
PEOIC	Percentage Eyes on Instrument Cluster
PEOR	Percentage Eyes on Road
PERCLOS	Percentage of Eye Closure
REML	Restricted Maximum Likelihood Estimation
RtI / TOR	Request to Intervene / Take-over request
SA	Situation Awareness
SAE	Society of Automotive Engineers
SAGAT	Situation Awareness Global Assessment Technique
SART	Situation Awareness Rating Technique
SD	Standard Deviation
SDLP	Standard Deviation of Lateral Position
SuRT	Surrogate Reference Task
SWT	Shapiro-Wilks Test
TOT	Take-over time
TTC	Time to collision

Contents

1 Introduction

The last years have seen two major trends in automotive research and innovation: electric cars and automation. While the number of electric cars available to consumers is rising every year, the development of automated driving in general is more erratic. Assistance systems like cruise control or parking sensors have been known for decades and were improved in evolutionary steps. Recently, the constant development of computational power and sensor technology has promoted a huge step in the development of commercial vehicle automation (Maurer, Gerdes, Lenz, & Winner, 2016, p. 90).

The *Society of Automotive Engineers* (SAE) (SAE J3016, 2018) defines six levels of automated driving, from Level 0 - manual driving to Level 5 - fully automated driving. Currently, Level 2 systems, or partially automated vehicles are already available to consumers. Level 3 systems, or conditionally automated vehicles are expected to be offered in the near future (Hafner, 2020). Consequently, past and current research both from industry and academia focused on the relevant safety and comfort issues related to the human factors of conditionally automated driving (CAD).

The six levels of automation must not be understood to represent an ordinal scale but rather a helpful framework to better distinguish between the levels of automated driving from a functional point of view. Level 2 systems take-over parts of the dynamic driving task (DDT) but still require the driver to monitor the system as integrative part of the driver-vehicle control loop (Bubb, Bengler, Grünen, & Vollrath, 2015, p. 28). CAD constitutes an automation controlling the DDT while the driver can potentially engage in non-driving related activities or tasks. He remains a receptive fallback level in case a system limit is met and the automation issues a request to intervene (RtI) (SAE J3016, 2018). Drivers then have to react to the RtI, reenter the driver-vehicle control loop and resolve the situation by driving manually. This process is referred to as take-over in CAD.

This take-over process represents a highly safety-relevant transition in addition to any consequences on liking, perceived comfort, acceptance and trust of the underlying automation. Therefore, the human factors research on CAD has focused on the take-over process. Empirical work on factors like the available time budget, traffic density or non-driving related tasks has already yielded valuable insight into the transitional process from CAD to manual driving. Any transitions to different levels of automation are possible and thought upon. The human factors research community especially centered on the transition from CAD to manual driving to allow a more transparent understanding of maximal human performance. The driver state in general, its development during automated driving and its effect on take-over performance still contain key questions on specific topics like prolonged automation duration or non-driving related tasks (NDRTs).

This thesis focuses on effects from a change in driver state on take-over performance, using modeling approaches including idiosyncratic effects to predict take-over performance and enhancing the human-machine interface (HMI) for CAD.

Chapter 2 offers the theoretical framework for the empirical studies. The state of the art in literature for human factors research on CAD, take-overs and driver state is presented. The main construct of ideas and their underlying reasoning is based on the literature review which also allows a critical classification of results. Chapter 3 provides a brief summary of central findings from the literature review, the derivation of the research questions and the composition of the empirical studies. Chapter 4 provides the general methodological basis for the studies and allows insight into the experimental setup. Identical parts within the individual method sections of the experiments are depicted in this section. Chapters 5 and 9, each based on one empirical study, present the individual method, results and discussion for the underlying research questions. These results have not been pre-published to this thesis and are analyzed in detail in this work. Chapters 6 and 8 contain a brief summary of the main findings from two additional experiments, that have been pre-published to this thesis in Radlmayr, Brüch, Schmidt, Solbeck, and Wehner (2018) and Radlmayr, Fischer, and Bengler (2019). Chapter 7 focuses on a modeling approach based on the data from Chapters 5 and 6. Chapter 10 offers limitations to the empirical work in general and puts the results, conclusions and discussion of all four experiments in critical perspective to existing research including a summary and an outlook on future topics and research.

2 Take-over performance in conditionally automated driving

This chapter provides the theoretical basis for this thesis. The relevant literature and state of the art on take-overs in CAD are presented. Focus is put on the driver state, its development during automated driving and potential effects on take-overs. Different effects on take-over performance are summarized and put into perspective to allow a critical discussion of results from this thesis.

When referring to automated driving in general and CAD in particular, the underlying definition is taken from the Society of Automotive Engineers (SAE) (SAE J3016, 2018). From a historical point of view the document is based on the definition of levels of automation by Gasser and Westhoff (2012). They suggested five levels, from the lowest one - manual driving to the highest level - full automation. The definition from the SAE adds one additional level and has the six levels ranging from Level 0 - manual driving or "No Driving Automation" to Level 5 - "Full Driving Automation". The levels differ in the categories "Dynamic Driving Task" (DDT), "Dynamic Driving Task Fallback" (DDT Fallback) and "Operational Design Domain" (ODD). Concerning the DDT, the levels differ in the way and by whom the longitudinal and lateral control of the vehicle is executed, either being the automation or the driver. In addition, the "Object and Event Detection and Response" (OEDR) is either executed by the system or the driver. The DDT Fallback is distinguished between system or driver. The various combinations of DDT, DDT Fallback and OEDR are matched with the ODD, where the design domain is either not available (in manual driving), limited (for Level 1 to 4) or unlimited (in Level 5) (SAE J3016, 2018). Figure 2.1 provides an overview on the six levels of driving automation and the variations of DDT including fallback, OEDR and ODD. Level 2 systems are already available for customers, e.g. from Tesla, BMW, Audi, Mercedes-Benz or Volvo. While the DDT is being executed by the automation, drivers are still responsible for monitoring the environment, traffic and the system, since the OEDR responsibility lies with them. Drivers must detect system failures or limits and take over control in case the system fails.

Level 3, CAD introduces a fundamental paradigm change (Lorenz, Hergeth, Kerschbaum, Gold, & Radlmayr, 2015). The OEDR is being executed by the system along with the longitudinal and lateral control. Drivers do not have to monitor or supervise the system but remain fallback-ready users (SAE J3016, 2018). Concerning the new role of drivers, they are potentially free to divert their attention to other tasks not related to driving or monitoring. In case the system encounters a system limit, the receptive user becomes the driver during fallback, he or she takes over (SAE J3016, 2018). The ODD is limited and using the system is restricted to e.g. highways or interstates. This leads to transition processes between the various levels of automation. CAD is the first level in which drivers are allowed to exit the driver-vehicle control loop (Bubb et al., 2015, p. 28). Therefore, the transition to lower levels of automated driving proposes new challenges to the safety and comfort of drivers. The OEDR lies with the system and drivers have to regain the OEDR responsibility. CAD in general is affecting the drivers' activity, attention processes, workload, situation awareness, behavioral adaptions, acceptance and trust (Navarro, 2018). The transition from CAD to lower levels and the accompanying paradigm change is of highest interest concerning the successful introduction of Level 3 systems. Research on factors affecting these take-overs has seen great focus in the last years.

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

Figure 2.1: Overview of the levels of driving automation from SAE J3016 (2018).

Known research and literature is presented and discussed in Chapter 2.2. This spike in research on take-overs in CAD was accompanied with an abundance of new concepts and definitions. While the discussion on some of them is ongoing, this section clarifies which definitions are used throughout this thesis, where they originate from and why they were chosen to serve as theoretical basis.

- **Conditional Driving Automation, conditionally automated driving (CAD)**
  CAD is referring to a Level 3 system following the definition of levels of automation of the SAE J3016 (2018). This can be confused with the corresponding level of the German Federal Highway Research Institute, or *Bundesanstalt für Straßenwesen (BASt)*, which can be literally translated to highly automated driving (Gasser & Westhoff, 2012). Following the SAE definition, a High Driving Automation incorporates the option of the system reaching a minimal risk condition, if an ODD limit is reached and there is no alternative DDT fallback. This thesis is focusing on Level 3 systems or CAD, in which a receptive fallback ready user responds to a request to intervene, performs fallback and resumes DDT performance (SAE J3016, 2018).

- **Take-over process**
  The transition process in which a driver performs as fallback and resumes the DDT is labeled to be a take-over (process). A take-over does not necessarily incorporate a transition from Level 3 to Level 0 without any assistance systems but could also be understood as transition from Level 3 to either Level 2 or 1. Most of the research on take-overs focuses on the transition from Level 3 to 0 since the take-over performance is best assessed without interference from assistance systems (Level 1) or partial driving automation (Level 2, PAD). Most of the literature presented in Chapter 2.2 on transitions from Level 3 to 0 use the term take-over process, e.g. see Damböck (2013).

- **Request to Intervene (RtI), Take-over request (TOR)**
  CAD implies drivers to respond to a Request to Intervene (RtI) or take-over request (SAE J3016, 2018). The term take-over request (TOR) was commonly used in publications up to 2017, e.g. see Damböck (2013), van den Beukel and van der Voort (2013), Gold (2016), Kerschbaum (2017). It was succeeded by RtI following a previous version of SAE J3016 (2018) and Marberger et al. (2017). Both documents are part of the ongoing revision of the ISO technical report ISO/TR21959-1:2018 which is in line with SAE J3016 (2018). Take-over request is synonymous with RtI. In this thesis, only RtI is used to comply with the ISO document.

- **Take-over situation**
  The empirical work in this thesis is based on evaluating take-over performance. The underlying premise requires a situation or scenario in which the transition process is assessed. The take-over situation is understood to be the testing situation that incorporates both chronological and spatial limits that define the beginning and end of a transition from CAD to manual driving. While there are many definitions on traffic situations or scenarios, there is no common ground in literature on what precisely is defined as a situation in general (Schneider, 2009).
  The take-over situations in this thesis are based on the concept defined by Gold, Naujoks, Radlmayr, Bellem, and Jarosch (2017). This concept is part of the results from the German nationally funded research project on cooperative, conditionally automated driving - Ko-HAF (*Kooperatives, hochautomatisiertes Fahren – Ko-HAF*, 2018). A take-over situation is defined, both chronologically and spatially, to start with the RtI and end with the system limit or reason for the RtI. More details on the specific situations in this thesis are offered in Chapter 4.

- **Take-over performance**
  Take-over performance is an umbrella term for various metrics quantifying driver behavior during a take-over. The definition is taken from Gold (2016). The term incorporates time aspects, such as the TOT and quality aspects such as accelerations or the time to collision (TTC). Generally speaking, a better take-over performance is associated with a reduction of the TOT and accelerations and an increase in the TTC. At this point, there are no specific thresholds to distinguish between a good or bad take-over.

- **Non-driving related activities or tasks (NDRA/Ts)**
  Prior to the introduction of CAD, research was focused on distraction (Bengler et al., 2014). Manual driving was state of the art and the introduction of assistance systems and new functions of vehicles such as air conditioning control or navigation

systems raised the question of whether these systems are a distraction from manual driving (e.g. Baumann, Rösler, & Krems, 2009).
Manual driving is split into three categories (Geiser, 1985). Primary tasks of the driver consist of navigating, maneuvering and stabilizing the vehicle in addition with recognizing relevant information from the surroundings. Secondary tasks are necessary for driving in general, but are independent of the execution of the DDT, e.g. operating the indicators or wipers. Tertiary tasks are linked to the comfort of drivers, such as climate control or controlling the radio. The labeling of secondary and tertiary with respect to the primary driving task raised questions of distraction, diversion of attention and multi-tasking abilities of drivers. A stringent application of a common understanding and methodology avoided negative side effects (Bengler et al., 2014).
Due to the inherent paradigm change of CAD, tasks that were previously labeled to be tertiary become new - non-driving related - primary tasks. CAD imposes a paradigm shift from a dual-task paradigm to a sequential-task paradigm (Lorenz et al., 2015). In order to differentiate new tasks and the new paradigm from the existing research on distraction, activities or tasks that are being executed during automated driving are labeled as non-driving related tasks (NDRTs). An overview on the abundance of new NDRTs during automated driving, known effects and their operationalization within this thesis' empirical work is provided in chapters 2.2 and 4.

- **Human-machine-interface (HMI)**
  The term HMI in an automotive context includes all parts available for drivers that convey information and/or are used for operation of the vehicle (Bubb et al., 2015, p. 272). Interfaces conveying information to the driver range from visual displays, to auditory or haptic signals. Control elements used for operation of the vehicle range from pedals, the steering wheel and any additional sensors, e.g. buttons, microphones, to additional controls used for engaging in NDRA/Ts (Bubb et al., 2015, p. 273). Focused on automated driving in general and CAD specifically, research and development of optimized HMIs is directed towards the full spectrum of HMIs e.g. conveying haptic information (Petermeijer, 2017) or transforming the steering wheel (Kerschbaum, 2017).
  The experiments in this thesis aim to optimize the visual interaction of drivers with NDRTs and the display of information during a take-over process. A head-up display (HUD) is utilized in two experiments in this thesis to optimize the visual interaction between system and drivers.

The definition of CAD highlights the importance of the take-over concerning the interaction of drivers with such systems. With the limited ODD of CAD, the transition to manual driving provides the highest understanding of the underlying human factors and offers the most leverage of evaluating novel HMI-concepts. The transition from CAD to manual driving is at the core of the empirical work in this thesis.

The field of research on automated driving in general has given rise to many questions, such as testing standards, liability details, security concerns or data privacy (Fagnant & Kockelman, 2015; Maurer et al., 2016). These questions and concerns are inherently connected to the successful introduction of automated driving. PAD systems are in use today whereas CAD is understood to be the next system available to users. From a functional perspective, vehicle control is executed by the automation for level 2 and 3. From a human factors perspective, the paradigm change of CAD imposes many new

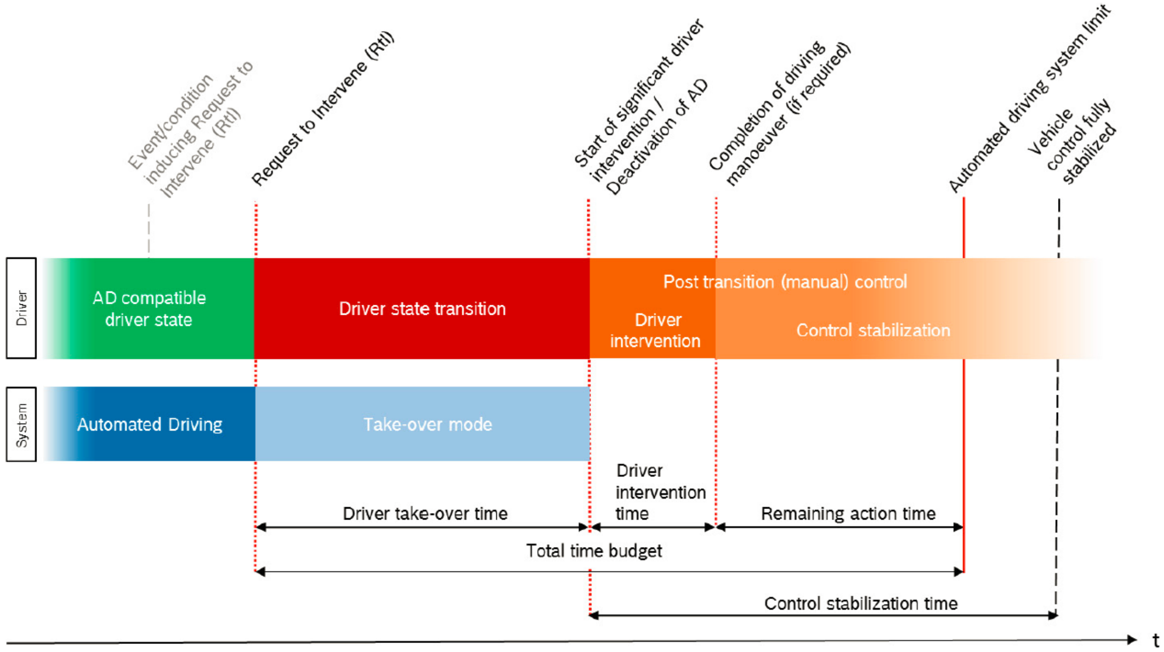


Figure 2.2: Graphical representation of the take-over process from Marberger et al. (2017).

questions due to drivers exiting and entering the driver-vehicle control loop. The take-over in CAD is identified to be the crucial element (Gold, 2016). The underlying understanding of take-overs in this thesis is depicted in Figure 2.2. The process integrates both sequential and parallel procedures and allows the definition of the driver availability for a take-over. The take-over mode in Figure 2.2 is disregarded for this thesis to allow a transparent analysis of the human performance without support or interaction from system inputs. Control of the DDT is labeled in blue in Figure 2.2. The definition underlines the necessity to predict the intervention time to allow an accurate understanding of the availability prior to a take-over and gives reason to the modeling approach in Chapter 7. The definition for driver availability is presented for the take-over time (TOT) (Marberger et al., 2017). While the definition allows an understanding and quantification for TOT, an integration of additional metrics of take-over performance, such as accelerations, can be applied to the concept of driver availability.

- **Driver availability**
  Driver availability for a take-over is defined as the fraction between time necessary for a driver intervention and the time budget for a take-over. In case the time necessary exceeds the time budget, drivers would be unavailable for a take-over. The definition as fraction allows the quantitative comparison between different levels of driver availability and is partly based on the general theory of driver behavior from Fuller (2005).

The take-over process depicted in Figure 2.2 and the assessment of driver reactions between the RtI and the system limit serve as empirical basis for this thesis. The following sections also feature results based on different frameworks or models to allow a critical demarcation of results and conclusions.

General information on driver assistance systems, which for many aspects of CAD form the basis of methods and approaches, can be found in Winner, Hakuli, Lotz, and Singer (2015). While the following sections provide an overview of relevant literature concerning empirical findings, some of the publications originate from specific projects that focused on human factors challenges of Level 2 and 3 automated driving.

The first research projects on automated driving such as Prometheus (Williams & Preston, 1987) demonstrated the technical possibilities in general, while focus was not put on human factors questions or the commercial applicability of such systems. Research on assistance systems, such as adaptive cruise control, can be understood as research on automated driving of Level 1 (SAE J3016, 2018). The dissertation from Gold (2016) provides a very thorough overview on the history of research on automated driving and the main internationally or nationally funded projects are listed. This paragraph focuses on projects that have been published or started recently. The project HAVEit (Flemisch et al., 2010) focused on Level 3 automated driving and affiliated questions such as driver monitoring and the necessary technology to allow the safe and comfortable introduction of automated driving. The take-over of human drivers at system limits of the vehicle automation was not the main focus at the time. The results from HAVEit were regarded for the selection of the key elements of the generic HMI used in this thesis and the necessity of using eye-based systems to assess the driver state during automated driving (Flemisch et al., 2010). The H-mode project and metaphor (H: horse) (Bengler & Flemisch, 2011; Flemisch, Bengler, Bubb, Winner, & Bruder, 2014) focused on optimizing the HMI between vehicle automation and driver by allowing a continuous arbitration of responsibility concerning vehicle control between a loose and tight reign. Haptic feedback was used to support the driver by continuous feedback and automating parts of the DDT without the option of drivers exiting the driver-vehicle control loop. The H-mode concept showed great benefits concerning the overall driving performance. The introduction of Level 3 and the associated possibility of engaging in NDRTs leads to drivers exiting the driver-vehicle control loop. This paradigm change was not integral part of H-mode and results from H-Mode are not regarded for this thesis. The project AdaptIVe (Langenberg, Bartels, & Etemad, 2014) provided human factors recommendations for the introduction of vehicle automation in general, but targeted functional requirements and decision strategies for collaborative automation (Kelsch et al., 2017). Collaborative automation differs compared to the sequential shift between manual and automated driving associated with CAD. The results from AdaptIVe underlined the need of assessing the driver state in the context of CAD. The project UR:BAN (Bengler, Drüke, Hoffmann, Manstetten, & Neukum, 2018) focused on assistance systems in more complex urban environments, giving way to HMI solutions regarding HUDs and novel display concepts. The results are detailed and integrated in the iterative development of the optimized HMI in Experiment 4 in Chapter 9.

The results from these projects motivated the project Ko-HAF which provided the framework and key theoretical understandings of this thesis. Additional driver models and frameworks of driver behavior are depicted in the next section.

2.1 Driver models and constructs of driver state

Both theoretical and quantitative driver models have been developed in the last decades to allow a prediction of driving behavior. A comprehensive overview of the fundamental link between automated driving and associated effects on driver behavior can be found in Bengler (2015). This includes well-known constructs and theories from the field of human factors not specifically aimed at automated driving. The models and constructs in this section differentiate between theoretical frameworks and quantitative models of human behavior. The review summarizes key models that were regarded in the development of the Ko-HAF framework and the principal understanding of driver behavior and modeling. The Ko-HAF framework serves as theoretical basis for the empirical work in this thesis and is aimed primarily at CAD.

The model human processor (Card, Moran, & Newell, 1986) explains and quantifies human performance from an engineering point of view in a dedicated environment. In the beginning of research on automated driving, the technical possibilities of realizing self-driving vehicles were of highest interest (Maurer & Dickmanns, 1997), while the human factors community focused on predicting manual driving performance (Sheridan, 1966). A spike in theoretical frameworks or qualitative links between different aspects of human behavior was reflected in Michon (1985) and later in Ranney (1994). The focus of these publications was manual driving and a critical view on proposed psychological behavior models. A general link between under- and overload conditions concerning arousal and performance was established by the Yerkes-Dodson law (Yerkes & Dodson, 1908). Both under- and overload conditions were associated with low performance while a medium arousal was understood to foster optimal performance. The law was generalized for the field of human factors in Teigen (1994). The connection between arousal and performance was fundamental basis for the design of experiments in Chapters 5 and 6 and the derivation of hypotheses concerning prolonged automated driving and the engagement in NDRTs. The multiple resources theory (Wickens & Liu, 1988) provides valuable insight into potential resource conflicts during the take-over. During the take-over process, the reallocation of resources to continue manual driving is competing with the current task or activity (Wickens, 2008). Depending on the modalities and specific characteristics of the NDRT such as effort to disengage, any engagement in NDRTs can be assessed theoretically. Potential effects range from showing negative consequences due to an overload of modalities up to helping drivers by preventing them from falling asleep due to underload conditions. Special focus is put on (visual) attention (Wickens & McCarley, 2007). While drivers are part of the driver-vehicle control loop during manual driving, attentional resources are distributed between the DDT and other, distracting tasks. The sequential switch between CAD and manual driving in a take-over in a short period of time accentuates the importance of attention. Misguided attention or attention on details not necessary for the take-over can greatly influence human performance. The necessity of monitoring attention as essential part of the driver state during automated driving can be derived. Following a holistic view on driver state, CAD as specific level of automated driving calls for driver monitoring (Rauch, Kaussner, Krüger, Boverie, & Flemisch, 2009; Müller & Bläsig, 2014). While the following models already focus on automated driving, information on the different challenges of manual driving are typically based on the distinction between behavior based on skills, rules and knowledge (Rasmussen, 1983).

The framework of situation awareness (SA) (Endsley, 1988) is contemplated specifically. In CAD, drivers are likely to loose SA when they exit the driver-vehicle control loop and are "out of the loop" (Kaber & Endsley, 1995). Since the OEDR responsibility lies with the system and drivers are allowed to engage in NDRTs, driver states including low or no SA are likely to occur. SA allows an intuitive explanation of take-over performance a posteriori. Based on the general concept of SA, a prediction of specific metrics of take-over performance is not feasible. However, the measurement of SA can be utilized by different methods such as the SAGAT (situation awareness global assessment technique) or the SART (situation awareness rating technique). The popularity of SA and its utilization within the research scope of take-overs in CAD is viewed critically in Chapter 8 and discussed incorporating the latest thoughts on SA in the context of automated driving (Endsley, 2015a, 2015b).

For specific psychological models of driving automation, Stanton and Young (2000) propose a model incorporating constructs such as the locus of control, mental workload or SA, but remain on a qualitative point of view. The necessity of integrating specific empirical findings in a unified model of driving automation is underlined to improve the outcome of design processes and the interaction of humans with automated vehicles (Stanton & Young, 2000). Recent years have seen a steep increase in the suggestion of both theoretical frameworks and quantitative models. Braunagel, Rosenstiel, and Kasneci (2017) suggest a model predicting take-over readiness with an accuracy of 79% integrating the complexity of the situation, the type of NDRT and the number of gazes to the road. The model provides an automated way of assessing driver readiness for a take-over and highlights the feasibility of eye-based measures to both assess changes of the driver state and to predict take-over performance. Inter- and intra-individual differences are not regarded to further increase prediction quality.

Two frameworks from Heikoop, de Winter, van Arem, and Stanton (2016) and Lu, Happee, Cabrall, Kyriakidis, and de Winter (2016) focus on the integration of relevant constructs of driver behavior. The interactions between e.g. attention and workload are discussed critically and qualitative links are established. The frameworks are not regarded for this thesis, since they lack the possibility of establishing quantitative relations as well. The framework developed in Ko-HAF features a theoretical classification of driver state components relevant to CAD and the take-over process as safety and comfort critical event. With the definition of driver availability, it also proposes a quantitative relationship for individual metrics of take-over performance.

The Ko-HAF framework is portrayed in Figure 2.3. Various factors, such as NDRTs, the RtI or the take-over situation affect the driver state. Key element of the framework is the transition process between a current and target driver state. The time needed for this transition process with respect to the available time budget of the situation allows the quantification of driver availability. Quantifying changes of the current driver state regarding e.g. the sensory or motoric state and analyzing the effect on take-over performance allows a precise evaluation of the new role of the driver in CAD. Motivational conditions, driver training and system experience are taken into account and the framework leads way to a comprehensive understanding of the take-over process in CAD. A more detailed summary of the construction and understanding of the concept of driver availability and the framework can be found in Marberger et al. (2017). In order to comply with the general wording used in existing literature and the scope of this thesis exceeding the current definition of driver availability by regarding more metrics than TOT, driver state is used throughout this thesis. Driver availability for a take-over can be understood to be a specific level of the general driver state in CAD.

For insight on individual differences in human-automation interaction, this thesis refers to the integral overview provided by Körber (2018). Additional constructs highly discussed and researched in the context of CAD such as trust, acceptance and workload go beyond the scope of this thesis and are not discussed in this literature overview. Well established ground truths of human automation interaction, such as the ironies of automation (Bainbridge, 1983) or challenges of appropriate automation design (Parasuraman & Riley, 1997) are not detailed in the literature overview, but serve as basic knowledge regarding the design of experiments and critical considerations of results in this thesis.

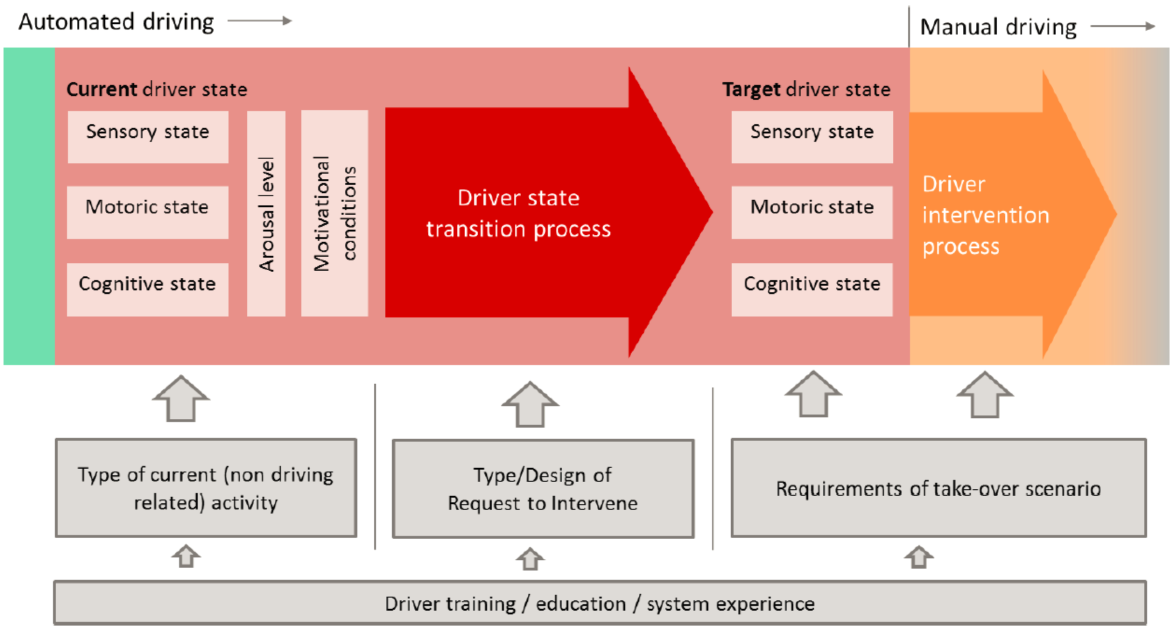


Figure 2.3: Framework of relevant factors of the take-over in CAD by Marberger et al. (2017).

2.2 Effects on take-over performance

This section focuses on publications and findings that put emphasis on various factors affecting take-over performance. The research questions detailed in Chapter 3 are derived from the following literature review. The focus on changes of the driver state and the resulting take-over performance in this thesis aims to answer open questions in the field of research.

The great increase in research and publications on take-overs in the last years was deduced in the dissertation from Gold (2016), where a modeling approach for the prediction of take-over performance was presented. Prediction quality was validated and compared to existing research from other institutes and companies. Findings emphasize the importance of considering changes in the driver state by utilizing driver monitoring. The predisposition of drivers also revealed significant model improvements and should be considered for future modeling attempts (Gold, 2016). The thesis from Körber (2018) highlights inter-individual differences in human-automation interaction and emphasizes the need for a quantitative evaluation of these differences. The comparison with other factors such as situational or state effects is essential (Berghöfer, Purucker, Naujoks, Wiedemann, & Marberger, 2018; Körber, 2018) to allow a feasible prioritization of measures to further optimize the take-over in CAD.

The increase in publications on take-over research also promoted meta-analyses and summaries of most important findings similar to this chapter. An overview of well-known effects on take-over performance such as different situations or NDRTs can be found in de Winter, Happee, Martens, and Stanton (2014) and Happee, Gold, Radlmayr, Hergeth, and Bengler (2017). The general approach of initial experiments on take-overs in CAD included comparing manual drivers to drivers taking over in the same situation (Damböck, 2013; Radlmayr, Gold, Lorenz, Farid, & Bengler, 2014; Ito, Takata, & Oosawa, 2016). Manual drivers show faster reaction times since they never exit the driver-vehicle control loop. Concerning quality measures such as TTCs or accelerations, minor differences, highly dependent on the specific situation, can be observed. The take-over performance in general resembles maneuvers from manual drivers concerning quality measures. Carryover effects can be observed (Skottke, Debus, Wang, & Huestegge, 2014), but remain small compared to situational influences.

The majority of publications focused on finding differences between groups or conditions when assessing take-over performance. Modeling approaches to predict behavior were not limited to Gold (2016), but can be found in e.g. Zeeb, Buchner, and Schrauf (2015) as well. Results underline the feasibility of predicting take-over performance to allow a potential future, integrated assessment of drivers being a fallback ready option or not. Results show, that motor readiness can be carried out reflexively, whereas cognitive processing of the take-over is impaired by e.g. drivers engaging in NDRTs (Zeeb, Buchner, & Schrauf, 2016).

Great focus is put on time aspects of the take-over process to not only find differences between effects but also to determine a maximal time in which participants can take-over safely. Studies on finding the minimal TOT were considering time budgets between four to eight seconds (Damböck, 2013; Gold, 2016) whereas more recent studies suggest a time budget of ten seconds or more for a complete take-over (Wan & Wu, 2018). Reaction times can be accelerated to a minor extent with smaller time budgets (Gold, 2016), but key element of the TOT is understanding the take-over situation before regaining manual control (McDonald et al., 2019). This also includes coping with effects from repeated take-overs, changing levels of automation and the driving environment calling for models accumulating all these effects (McDonald et al., 2019). Age, potentially increasing reaction times of the take-over, showed no effect for time aspects but led to stronger brake maneuvers compared to young drivers (Körber, Gold, Lechner, & Bengler, 2016a). Middle-aged drivers, selected due to their increased experience in manual driving on a daily and annually basis, showed faster reactions compared to younger drivers (Wright, Samuel, Borowsky, Zilberstein, & Fisher, 2016).

The process of regaining an understanding of the take-over situation, re-entering the driver-vehicle control loop and driving manually requires a high amount of visual attention. The visual behavior prior to and during the take-over process is of utmost importance for the successful execution of the take-over process (Louw, 2017; Vlakveld, van Nes, de Bruin, Vissers, & van der Kroft, 2018). Eye-tracking is of high interest to allow a detailed understanding of effects on the take-over process and to allow a future modeling attempt of take-over performance (McDonald et al., 2019).

A comprehensive overview on determinants of TOT can be found in Zhang, de Winter, Varotto, Happee, and Martens (2019). The meta-analysis focused on factors influencing the TOT. Results revealed that the following factors can shorten the TOT: a high urgency of the situation, not using a handheld device, not performing a visual NDRT, having experienced another take-over scenario before in the experiment, and experiencing an auditory or vibrotactile RtI as compared to a visual-only RtI or no RtI. A familiarization with CAD based on measurements of seating position, glances off the road, NDRT engagement and self-reports can be achieved after ten minutes and is correlated with gender and previous experience with advanced driver assistance systems (Omozik, Yang, Kuntermann, Hergeth, & Bengler, 2019).

In conclusion, effects on the take-over performance were found for situational factors, different NDRTs, the RtI or the HMI in general and the driver state. Individual differences that can be assessed using e.g. eye-tracking and predispositions such as age, gender and previous experience should be regarded for a modeling approach. The following sections detail the listed factors and broaden the literature review to allow a specific derivation of research questions. Overall literature findings were regarded for the design of experiments of the four studies in Chapters 5, 6, 8 and 9 and the modeling of take-over performance in Chapter 7.

2.2.1 The take-over situation

Every take-over is associated with a specific traffic situation in which drivers have to regain control. While the system limit accounting for the RtI is typically static, the take-over situation is dynamic. The definition and classification of take-over situations used in this work is based on Gold et al. (2017) and is addressed in more detail in Chapter 4.

Situational factors affecting take-over performance have been reported for almost all experiments regarding different situations. Time budget was addressed by Damböck, Farid, Tönert, and Bengler (2012) and Gold, Damböck, Lorenz, and Bengler (2013) and future experiments incorporating these findings. The first time budget that was reported to be sufficient for potentially all situations was published to be ten seconds (Melcher, Rauh, Diederichs, Widlroither, & Bauer, 2015). More recent publications on take-over performance in different situations consistently reported that environmental factors and the level of automation are the most prevalent factors (R. C. Gonçalves, Quaresma, & Rodrigues, 2017). While the time budget highly affects the situational criticality by moderating the time available for a take-over, traffic density was also identified to exert a substantial effect on take-over performance. This effect has been also reported for manual driving (Baldwin & Coyne, 2003) and significantly increases TOT and worsens quality measures in CAD (Radlmayr et al., 2014; Gold, Körber, Lechner, & Bengler, 2016). A general increase in criticality of take-over situations can lead to a decrease in TOT to a minor extent (Roche & Brandenburg, 2018), but is accompanied with worse take-over quality (Gold, 2016). Varying traffic conditions in CAD led to behavioral changes with drivers directing more attention to the road while experiencing higher traffic densities (Jamson, Merat, Carsten, & Lai, 2013). This effect has to be discussed critically since drivers do not have to monitor in CAD but are allowed to engage in NDRTs.

More recent research focused on the time needed to recollect a detailed depiction of surrounding traffic in take-over situations and reported seven to twelve seconds to be sufficient for a spatial representation of the situation (Lu, Coster, & de Winter, 2017). In conclusion, the effect of situational factors on take-over performance such as time budget or traffic density appears to be of highest magnitude compared to other factors. Different take-over situations should be considered to allow a more comprehensive understanding of resulting take-over performance measures.

2.2.2 Non-driving related tasks

The abundance of potential NDRTs drivers might engage in during CAD, highlights the complexity and scope of answering research questions on effects from different NDRTs on take-over performance. NDRTs are at the core of the paradigm change between Level 2 and Level 3 automation (Lorenz et al., 2015). The DDT is executed by the automation and the engagement in NDRTs represents a major gain in time utilization for drivers. Drivers are allowed to engage into a variety of activities or tasks and show an interest in doing so (Pfleging, Rang, & Broy, 2016). Drivers are both interested in activities known from manual driving, e.g. listening to music and NDRTs only available in CAD, e.g. writing text messages or browsing the Internet (Pfleging et al., 2016). This section offers a small overview on recent findings.

Drivers in CAD are not obliged to engage in NDRTs but are free to monitor the automation or the surroundings or simply not engage in any specific activity or task at all. In case drivers take up the possibility, an engagement in NDRTs leads to more challenges when regaining manual control during a take-over (Louw, Merat, & Jamson, 2015; Bueno et al., 2016). These results focus on the effect of NDRTs on take-over performance compared to no engagement in NDRTs for short periods of automated driving. Recent studies revealed a more complex structure of effects, sometimes failing to find clear and systematic differences between experiments (Jarosch, Gold, et al., 2019).

Engagement in NDRTs can be instructed for manipulation purposes in lab conditions but voluntary engagement, prompting motivational aspects, showed no impairment on take-over performance (H. Clark & Feng, 2017). While Radlmayr et al. (2014) showed similar take-over performance comparing a visual-motoric and a cognitive task, participants in Gold, Berisha, and Bengler (2015) showed degraded performance for visual-motoric tasks in well practiced and known situations. In both experiments, participants were instructed to engage in NDRTs and results show the interaction between the effect of situational factors and NDRTs. In cognitive demanding situations, cognitive NDRTs show the same decrement on take-over performance compared to visual-motoric tasks. In well-practiced and less cognitively demanding situations, only visual-motoric NDRTs show negative effects on the take-over performance (Gold et al., 2015).

Concerning specific properties of NDRTs, occupation of one or both hands (Naujoks, Purucker, Wiedemann, & Marberger, 2019) and the steps needed to disengage from the NDRT (J. Gonçalves & Bengler, 2016) exert a negative effect on take-over performance by increasing TOT and worsening specific take-over quality metrics, such as lateral accelerations. Hand-held devices, which occupy the hands and need to be put away before a manual intervention are frequently identified to increase TOT. Situations that are less critical increase the magnitude of this effect (Jarosch, Gold, et al., 2019). A longer engagement in NDRTs or more experience with NDRTs in vehicles capable of CAD could foster additional problems, such as new postures leading to a decrement of take-over performance (Yang, Klinkner, & Bengler, 2019). NDRTs motivating "out of driving" postures show a critical prospect of affecting take-over performance because drivers have to relocate themselves (Jarosch, Gold, et al., 2019).

The most critical aspect of NDRTs affecting take-overs ties into the perception of the RtI. During CAD, drivers are fallback-ready users that need to take-over in case a system limit is met (SAE J3016, 2018). NDRTs preventing the perception of the RtI can cause severe delays or complete omissions of the take-over (Jarosch, Gold, et al., 2019).

Self-regulation, both for NDRTs and take-overs exerts beneficial effects on take-over performance by decreasing TOT and improving take-over quality (Ko & Ji, 2018; Eriksson & Stanton, 2017). These findings link to research on the effect of prolonged durations of automated driving and associated problems, such as vigilance decrements or an increase in drowsiness. Section 2.2.3 focuses on known effects from an increase of drowsiness, while this section highlights positive aspects of engaging in NDRTs to counter drowsiness or vigilance decrements during CAD. Drivers engaging in reading or watching videos are less likely to exhibit behaviors indicative of drowsiness compared to monitoring the automation (Miller et al., 2015). Studies associated with Ko-HAF revealed the potential of NDRTs to counter an onset of drowsiness or fatigue (Jarosch, Kuhnt, Paradies, & Bengler, 2017). Driver drowsiness can be managed up to a certain extent by offering a targeted use of different NDRTs (Weinbeer, Muhr, & Bengler, 2019).

The most prominent findings from literature in addition with recent results from empirical studies were accumulated in key messages in the project Ko-HAF (*Kooperatives, hochautomatisiertes Fahren – Ko-HAF*, 2018). The following attributes of NDRTs were found to increase TOT compared to not engaging in NDRTs

- Holding an object (e.g. a mobile device) in one/both hands
  Manual interaction (one/both hands) with mobile electronic devices
- Unusually strong rotations (>90°) of the torso
- Increased effort or multiple steps needed to fully disengage from a NDRT

whereas the following attributes showed no consistent effects:

- Visual or visual-motoric tasks (e.g. watching video, reading, texting) without occupation of one/both hands
- Cognitively demanding NDRTs affecting the cognitive transition

Generally, strong inter-individual differences were found concerning how NDRTs affect the driver state and the magnitude of associated effects depends on methodological issues like testing environment and situations, e.g. identical NDRTs lead to significantly different results if investigated in less critical and urgent situations in real traffic compared to driving simulators (Jarosch, Gold, et al., 2019).

They key messages from Ko-HAF on NDRTs include specific recommendations concerning research:

- Natural behavior, self regulation and motivational aspects of NDRTs must be considered in the experimental design.
- A safety assessment of NDRTs with respect to their effects on take-over performance can only be carried out taking into account the details/parameters/aspects of the test scenario (e.g. the available time budget).
- A „NDRT lockout“ simultaneously with the RtI can speed up the driver response to the RtI. A lockout is a system-initiated interruption of the NDRT performed on the vehicle-integrated infotainment system or on connected portable devices with an additional presentation of the RtI on the respective screen.

In addition to showing moderate benefits for take-over behavior, a NDRT lockout is highly accepted by drivers (Wandtner, Schömig, & Schmidt, 2018).

Concerning a classification and utilization of NDRTs for their application in research based on previous findings, two major characteristics of NDRTs can be differentiated. Realistic NDRTs can be evaluated to directly quantify the effect from specific NDRTs on take-over performance. Standardized NDRTs with a focus on specific parameters, such as modality or steps to en-/disengage, can be utilized to deduce more general findings. Ko-HAF classified and ordered various NDRTs to allow an easy decision making process in the design of experiments on NDRTs and their effect on take-overs (*Kooperatives, hochautomatisiertes Fahren – Ko-HAF*, 2018). The resulting catalog of NDRTs (Naujoks, Purucker, & Neukum, 2017) served as basis for the decision to integrate only standardized NDRTs in the experiments described in this thesis.

Concluding the overview on NDRTs, standardized NDRTs were deemed more feasible to allow an assessment of driver state changes - due to NDRTs - and their effects on take-over performance. The research questions in this thesis also focused on the interaction of NDRTs with respective take-over situations and most relevant issues linked to attention. A more detailed look on the standardized tasks used in this thesis is provided in Chapter 4.

2.2.3 Sleepiness, drowsiness, fatigue and vigilance

The effects from sleepiness, drowsiness and fatigue are regarded due to their potentially significant role in CAD. Drivers could use the time during active CAD to relax or daydream promoting vigilance decrements or an onset of drowsiness. The conclusion on NDRTs revealed that engaging in NDRTs can have positive effects on driver state regarding the arousal level by countering an onset of drowsiness. Thus, prolonged, monotonous periods of CAD or the desire from drivers to rest is likely leading to an increase of drowsiness causing drivers potentially falling asleep.

Prior to the summary of relevant findings in literature, the definition of terms is addressed. While the terms sleepiness, drowsiness and fatigue are often used simultaneously in relevant literature on CAD, past definitions distinguish between the terms. A comprehensive overview of different definitions, associated effects, relevant findings and most recent research can be found in Radlmayr, Feldhütter, et al. (2018). The definition of these terms is based on Johns (2000) and Johns (2007) and distinguishes drowsiness and sleepiness from fatigue. Drowsiness describes the interval between being awake and asleep and is synonymous with sleepiness (Johns, 2007). Fatigue represents changes of the driver state stemming from prolonged engagement in strenuous tasks or activities. Fatigue can be relieved by rest, whereas drowsiness can be mitigated by sleep (Radlmayr, Feldhütter, et al., 2018). This is in contrast to e.g. May and Baldwin (2009), where fatigue contains both task-related fatigue and sleep-related fatigue where the latter represents drowsiness/sleepiness in Johns (2007). For this thesis, only the term drowsiness is used to address driver state changes originating from prolonged, monotonous periods of automated driving or underload conditions.

While visual attention in CAD was identified to be of high interest for the analysis of the take-over process in Section 2.2 and is analyzed in the empirical studies in this thesis, vigilance is not regarded. Vigilance is synonymous with sustained attention and describes a state of readiness to detect and to react to stimuli, which appear at random and seldom intervals for an extended time (Körber, Schneider, & Zimmermann, 2015). The sustained ability to focus on a specific task is typically researched for long, monotonous periods of manual driving or PAD. Drivers in PAD are required to monitor the automation and the traffic situation at all times and vigilance decrements play a crucial role in evaluating the success of monitoring. The paradigm change for CAD does not require drivers to monitor the system and a specific focus on vigilance is not regarded in this thesis. Potential effects from prolonged, monotonous automated driving are understood to cause an increase in drowsiness, since drivers are not trying to monitor the automation unless instructed otherwise. Potential problems from vigilance decrements as first step towards an onset and development of drowsiness are regarded in Körber, Cingel, Zimmermann, and Bengler (2015). For further information on attention and vigilance in general, and vigilance as potential preliminary state prior to becoming drowsy, Kahneman (1973), Warm, Parasuraman, and Matthews (2008) and Parasuraman and Manzey (2010) are recommended.

Research on the effects from drowsiness on manual driving has been conducted for decades but typically relied on assessing driver inputs, such as the steering behavior (Daza et al., 2011). The results from analyzing driver inputs provide an accurate estimation of drowsiness levels but are not feasible for evaluating drowsiness in CAD. The disengagement of drivers from the DDT during active CAD inherently requires new methods that are not relying on driver inputs. Future application of drowsiness detection in vehicles featuring CAD should be non-intrusive.

Throughout the literature, a prolonged duration of automated driving is associated with an increase in drowsiness (J. Gonçalves, Happee, & Bengler, 2016). One hour of CAD compared to ten minutes of CAD led to significantly increased levels of drowsiness (Bourrelly et al., 2019). Empirical studies evaluated the development of drowsiness by either considering fixed-time approaches or issuing the RtI dependent on a specific level of drowsiness. Fixed-time approaches showed no increase in TOT but reported small changes in quality aspects of the take-over, such as higher lateral accelerations or a larger amount of participants braking to a full stop (J. Gonçalves et al., 2016; Feldhütter, Gold, Schneider, & Bengler, 2016). High individual differences in the onset and development of drowsiness led to approaches integrating a fixed level of drowsiness before the RtI was issued, independent of the time it took participants to develop that level (Radlmayr, Feldhütter, et al., 2018).

Resulting take-over performances after a fixed level of drowsiness did not show consistent significant changes concerning TOT (Weinbeer et al., 2017), but revealed contradicting findings. Minor changes concerning quality aspects of take-over performance were reported (Kreuzmair, Gold, & Meyer, 2017; Feldhütter, Kroll, & Bengler, 2018) in addition to no effects on both time and quality aspects (J. Schmidt, Stolzmann, & Karrer-Gauß, 2016; J. Schmidt, Dreißig, Stolzmann, & Rötting, 2017). The interpretation of overall results exposed high individual differences in the development of drowsiness with little to no effect on take-over performance (Radlmayr, Feldhütter, et al., 2018). Regardless, extreme levels of drowsiness should be abstained to avoid critical take-over performances in more critical situations (Radlmayr, Feldhütter, et al., 2018).

Typical time budgets (five to ten seconds) associated with CAD prohibit a safe wake-up process in case drivers fall asleep. A high level of drowsiness was identified as state prior to falling asleep. Thus, strategies to counter higher levels of drowsiness were also evaluated. Thermal stimulation to adjust the drowsiness level of drivers during automated driving was revealed as feasible counter measure (E. Schmidt, Ochs, Decke, & Bullinger, 2017). Weinbeer, Bill, Baur, and Bengler (2018) showed that a targeted suggestion to engage in NDRTs to postpone the further development of drowsiness was most widely supported.

The development of drowsiness can be moderated by the introduction of NDRTs. NDRTs either exert beneficial effects on drivers or lead to the development of fatigue due to excessive engagement in NDRTs. Various empirical studies analyzed the combination of NDRTs, their effect on drowsiness and potential consequences on the take-over performance (Naujoks, Höfling, Purucker, & Zeeb, 2018; Jarosch, Bellem, & Bengler, 2019; Jarosch & Bengler, 2019). In case the NDRTs were monotonous and strenuous, results did not show differences for the TOT but minor effects on quality aspects such as a larger amount of drivers braking to a full stop.

Including the key findings from Ko-HAF, it can be concluded that during CAD, drowsiness can develop or be induced quickly but might be subject to volatile changes. Drivers show strong inter- and intra-individual differences in the development of drowsiness. An increase of drowsiness under monotonous conditions could already be detected during shorter, uninterrupted automated drives (20 to 30 minutes). Under these conditions no significant effect on TOT and quality could be detected. Also, in longer, uninterrupted automated drives (up to 90 minutes) clear and consistent effects on take-over behavior could not be found (*Kooperatives, hochautomatisiertes Fahren – Ko-HAF*, 2018). Nonetheless, the assessment of drowsiness in CAD is of high interest to avoid the unacceptable state of sleeping drivers and to allow the precise utilization of methods countering higher levels of drowsiness as preliminary stage.

2.2.4 The human-machine interface

The HMI during automated driving and for the take-over process includes all elements transporting information between the vehicle and the driver (Bubb et al., 2015, p. 272). Hence, an optimization or targeted development of the HMI for CAD shows great potential to affect the driver state and take-over performance. The following overview on effects from different HMIs focuses on the take-over process. Both safety and comfort aspects of CAD highly depend on the appropriate design of the HMI.

Research and development on suitable HMIs for CAD or automated driving in general has seen a steep increase from both academia and car manufacturers in recent years (Kerschbaum, 2017). The RtI is fundamentally linked to perception and the possibility to convey information on the pending take-over situation. A visual display of the RtI at a moment when drivers are engaged in visual NDRTs on different screens could result in perception problems potentially leading to slower reactions or missed take-overs. Since CAD implies drivers to be fallback-ready, perception issues are critical.

Roche, Somieski, and Brandenburg (2019) found auditory RtIs leading to smaller TOTs and higher TTCs compared to visual-auditory RtI in combination with an improvement in experience for only auditory RtI. It was argued, that the visual component of a RtI could compete with the take-over situation itself for visual attention. Contrary, the key messages from Ko-HAF recommend the RtI to be multi-modal in order to unequivocally convey the necessity for taking over vehicle control (*Kooperatives, hochautomatisiertes Fahren – Ko-HAF*, 2018). In addition, the enrichment of generic HMIs by including speech outputs shortened "information processing"-reaction times, e.g. termination of NDRTs whereas "allocation of attention"-reaction times, e.g. first glance ahead, did not show significant effects (Forster, Naujoks, Neukum, & Huestegge, 2017). Subjective ratings favored additional speech output (Forster et al., 2017).

The potential to convey additional, semantic information using a haptic interface in the seat was examined by Petermeijer, de Winter, and Bengler (2015), Petermeijer (2017) and Cohen-Lazry and Katzman (2018). While various patterns and vibrations were analyzed concerning their effect on take-over performance, results showed the biggest influence on the perception of the RtI. Conveying more complex information, such as spatial direction information did not result in better take-over performance (Petermeijer, 2017). Perception of the RtI benefited from adding an additional modality compared to the typical visual-auditory HMIs (Petermeijer, 2017).

Most of the publications regarded the physical workplace of the driver to be static. Kerschbaum, Lorenz, and Bengler (2014) evaluated the decoupling of the steering wheel with additional work on a transforming steering wheel (Kerschbaum, Lorenz, & Bengler, 2015). Results showed that the decoupling or transformation of the steering wheel was rated usable (Kerschbaum et al., 2015). A physically transforming driver workplace facilitates the changing role of the driver in CAD without introducing new challenges for the safety-critical take-over process (Kerschbaum, 2017; Borojeni, Wallbaum, Heuten, & Boll, 2017).

A key improvement of HMIs for CAD targeted the provision of additional information, typically utilizing visual displays. The supplementary information leads to a more elaborated RtI design integrating not only a shortly timed stimulus but multiple steps of re-integrating the driver in the driver-vehicle control loop. A multi-step RtI was shown to increase drivers' situational awareness (Epple, Roche, & Brandenburg, 2018) and accelerate the disengagement from NDRTs and by this the TOT (*Kooperatives, hochautomatisiertes Fahren – Ko-HAF*, 2018).

During CAD, drivers request information on the system status and the transparency and comprehensibility of system actions (Beggiato et al., 2015). The range of information density and utilized methods ranges from incorporating contact-analogue HUDs augmenting the take-over situation in the windshield (Lorenz, Kerschbaum, & Schumann, 2014) to using ambient displays to communicate the system status (Borojeni, Chuang, Heuten, & Boll, 2016; Yang, Karakaya, Dominioni, Kawabe, & Bengler, 2018). While the technical application of augmented reality HUDs is still in the future, results showed that supporting the decision making process by suggesting potential corridors for manual driving during the take-over in the HUD is beneficial concerning the reaction type (Lorenz et al., 2014; Eriksson et al., 2019).

Research on newly developed HMIs for CAD also revealed an overall strong effect of the specific take-over situation at hand which should be considered to adapt the RtI accordingly (Bazilinskyy, Petermeijer, Petrovych, Dodou, & de Winter, 2018). Priming drivers with an appropriate maneuver (steering vs. braking) depending on the take-over situation, in case the information is available, decreased TOTs and increased TTCs (Borojeni, Weber, Heuten, & Boll, 2018). Using multi-modal solutions including haptic feedback via the steering wheel for a cooperative take-over underlined the feasibility of priming drivers for different situations (Kalb, Streit, & Bengler, 2018). Relying on the availability of information on upcoming situations in addition with a measurement of driver state, Winkler, Kazazi, and Vollrath (2018) suggested to adapt the warning strategy to both the situation as well as to the driver state.

A cooperative, multi-modal interaction concept targeted at developing the take-over process towards cooperative interaction revealed profits not only for the human-machine interaction within the vehicle but also the interaction between different road users (Zimmermann & Bengler, 2013). Motivating cooperative interaction between road-user by introducing time and a new currency improved the efficiency of the traffic system (Zimmermann & Bengler, 2013). Tactical level input (touchscreen, gesture, and haptic interfaces enabling bilateral driver-vehicle interaction) could reduce driver workload, reaction times, and improve driver behavior in addition to be highly preferred by drivers over a manual take-over (Manawadu et al., 2018).

Most of the published findings are based on the assumption of drivers being either instructed carefully or having had the chance to familiarize themselves with the system and thus the HMI. While a prior familiarization with RtIs showed positive effects on both take-over performance and trust in automation (Hergeth, Lorenz, & Krems, 2017), this cannot be assumed for all future applications of CAD. Users not knowing how to operate the driving automation system are likely to refer to a trial and error basis to establish experience. In order to attenuate detrimental effects observed in first contact human-automation interaction, the HMI has to be designed carefully to both support first contact and experienced users (Forster, Hergeth, Naujoks, Krems, & Keinath, 2019).

Concluding this section, general requirements for newly developed HMIs for CAD can be derived. The take-over process as key element of CAD regarding safety and comfort aspects should rely on a transparent, coordinated protocol (J. Clark, Stanton, & Revell, 2017). Information during CAD and the take-over, including the RtI, should be conveyed to drivers in a clearly perceptible, multi-modal way (*Kooperatives, hochautomatisiertes Fahren – Ko-HAF*, 2018). If available, continuous assistance prior, during and after the take-over can help with increasing safety and comfort margins of CAD (J. Clark et al., 2017). While these requirements are a first step towards a unification of HMI requirements, the future application of CAD in commercially available vehicles calls for a verification process and guidelines globally accepted to promote a successful introduction. A process for vehicle HMI verification including empirical results demonstrated the necessity for verified guidelines to aid the development of new HMIs for CAD (Naujoks, Wiedemann, Schömig, Hergeth, & Keinath, 2019; Forster, Hergeth, Naujoks, Krems, & Keinath, 2020).

2.3 Quantifying the driver state

The findings summarized in the previous sections allow a comprehensive understanding of most relevant effects on the driver state during automated driving and the take-over in CAD. This section details the most important methods and measures to quantify changes of the driver state. An extensive and thorough analysis and review of driver state monitoring system in the context of CAD are offered in J. Gonçalves and Bengler (2015), Cabrall et al. (2016) and Hecht et al. (2019). It can be derived that any future driver state monitoring technology must consist of non-intrusive systems and methods. Physiological measures, such as the skin conductivity or using an EEG can be used for a more detailed look or validation of other methods and measures but are not applicable in future vehicles.

The most prominent method to detect driver state changes, e.g. an onset of drowsiness, is eye-tracking (Chang & Chen, 2014). While head-based solutions typically offer a higher quality of tracking, these solutions are not qualified for future use in commercial vehicles (Radlmayr, Feldhütter, et al., 2018). Both pupil and eye-lid based measures are used to quantify changes in the arousal level of drivers. Eye-lid-based metrics were found to be most valid for detecting drowsiness (Radlmayr, Feldhütter, et al., 2018).

Eye-tracking systems are highly depended on the quality of tracking to allow a detailed and precise analysis. While all manufacturers of eye-tracking systems claim high numbers of successful tracking, hands-on experience in simulators and real driving environments showed typical tracking rates to range between 70% - 90%. The ISO-Norm 15007 (ISO/TS 15007, 2014) provides recommendations to calibrate and use eye-tracking systems to ensure tracking quality and identifies tracking rates above 85% to be of good quality and tracking rates of 95% to be excellent.

Applications of eye-tracking to detect driver state changes can be found in Feldhütter, Feierle, Kalb, and Bengler (2018) where a data fusion approach incorporates both eye- and head-tracking to detect drowsiness in the context of CAD. High validity and sensitivity rates give way to using a data fusion approach to improve the detection of relevant state changes. Challenges arise from a parallel engagement in NDRTs or posture adjustment in the seat due to the eyes shifting out of an optimal tracking area or space (Feldhütter, Hecht, Kalb, & Bengler, 2018). The percentage of eye closure in a specific time period (PERCLOS) (Dinges & Grace, 1998) was found to be highly reliable and often used to assess drowsiness (Darshana, Fernando, Jayawardena, Wickramanayake, & DeSilva, 2014). Most of the published papers on PERCLOS report additional measures to support PERCLOS, such as mouth shape and head position (Qiong, Jingyu, Mingwu, & Yujie, 2006), head movement as moderate indicator (E. Schmidt, Decke, & Rasshofer, 2016) or blinks and saccades (Schleicher, Galley, Briest, & Galley, 2008).

An optimal detection of drowsiness can be achieved by combining six eye-tracking parameters, percent eye closure (PERCLOS), eye closure duration, blink frequency, nodding frequency, face position, and fixed gaze (Bergasa, Nuevo, Sotelo, Barea, & Lopez, 2006). In addition, blink latency and gaze variability were reported to be valid measures for increased workload, e.g. due to the introduction of NDRTs. These findings were regarded to measure the driver state during automated driving to allow a thorough assessment of changes arising from e.g. prolonged durations of automated driving or the engagement in NDRTs.

While drivers could have their eyes closed due to a variety of reasons, validation methods such as self- or expert ratings of eye-based measures were utilized to allow an understanding if drivers were actually drowsy. The feasibility of determining a ground truth to assess driver state changes, especially concerning drowsiness, was demonstrated applying an expert rating (Weinbeer et al., 2018). The Karolinska sleepiness scale (KSS) as self reported measurement of drowsiness was also used to validate objective measures of drowsiness (Barua, Ahmed, Ahlström, & Begum, 2019).

Seat pressure mats were identified as a potential new, non-intrusive method to assess driver state changes. While the application of seat pressure mats is common in assessing seating comfort (Ulherr, 2019), driver state changes in the seat due to movement or posture changes due to the engagement in NDRTs are not assessed so far. Zilberg, Xu, Burton, Karrar, and Lal (2009) reported the seat movement magnitude to increase the detection of drowsiness in combination with using an intrusive method such as an EEG.

In summary, the most promising method of assessing driver state changes, including the onset of drowsiness or the allocation of attention, is eye-tracking. To allow an exhaustive assessment of driver state changes and their effect on take-over performance, additional sensors or a data fusion approach should be applied.

3 Research questions, contribution and structure of this thesis

The literature review in Chapter 2 provides a comprehensive overview and allows the identification of research gaps. The main conclusions from the literature review are repeated here to give way to the derivation of research questions.

The paradigm change of CAD in combination with a limited ODD incorporates the necessity to analyze and understand the development of the driver state. The driver state can be argued to change significantly since drivers temporarily leave the driver-vehicle control loop. While this new development is of high interest, key questions arise from the drivers being the fallback in case of a take-over. Both the engagement in NDRTs and the possibility of prolonged periods of active CAD could affect the arousal of drivers and thus, the take-over performance. The following research questions ensue:

1. How do prolonged periods of monotonous, automated driving affect the driver state?
2. How do different NDRTs affect the driver state?

These questions are targeted at periods of active CAD. The literature review also revealed significant effects on human performance from different take-over situations depending on their overall criticality or complexity. Potential effects on take-over performance from changes in the driver state should be matched with take-overs in different situations to allow a comprehensive understanding of underlying processes. This gives way to the next research questions:

3. How do potential driver state changes affect the take-over performance?
4. How does the effect from driver state changes on take-over performance compare to the effect of different situations?

Findings from previous modeling efforts revealed significant model improvements by considering the predisposition and inter-individual changes of drivers. Several recorded take-overs together with data from driver monitoring should be considered to answer the next research question:

5. How do effects from e.g. driver state changes, situational factors and NDRTs on take-over performance compare to the effect of predisposition and individual differences of drivers?

The aforementioned research questions are addressed in two experiments and a succeeding modeling approach in Chapters 5, 6 and 7. Based on these findings and the literature review in Chapter 2, the need for an optimization of the HMI for CAD is derived to increase safety and comfort margins of the take-over. Visual attention and addressing the strong effect of different take-over situations are at the center of derived research questions utilizing the HUD:

6. How can the HMI for CAD and the take-over be optimized by enabling the possibility of peripheral monitoring while engaging in NDRTs during CAD?

7. How can the HMI be optimized by offering additional information on the specific situation during the take-over process?

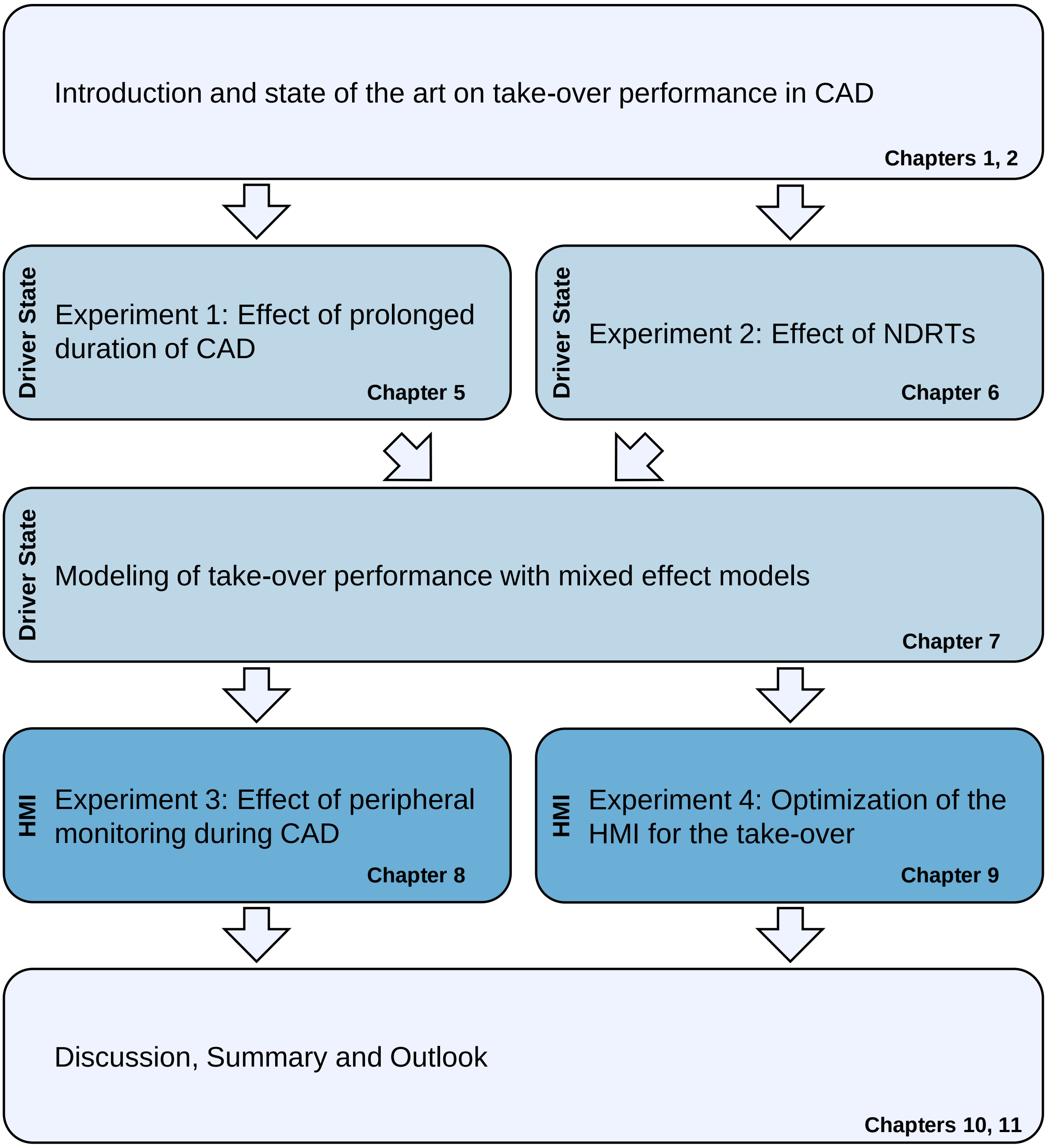


Figure 3.1: Overview of the structure of this thesis. Comprehensive chapters, empirical work on the effect of driver state changes and the experiments on the optimization of the HMI are colored differently to allow a better understanding. The modeling of take-over performance is based on data from Experiments 1 and 2.

The research questions six and seven are addressed in two experiments in Chapters 8 and 9. Figure 3.1 provides an overview of the structure of this thesis incorporating the empirical effort consisting of 4 experiments and the modeling approach. The research questions are addressed in individual depth in the respective experiments and the modeling approach and are specified in more detail below. The empirical work can be divided in two major areas of CAD. Experiments 1 and 2 and the modeling approach aim at driver state changes and potential effects on take-over performance. Experiments 3 and 4 target the optimization of the HMI to improve the take-over process.

Driver State

Experiment 1: Effect of prolonged duration of CAD

RQ1 How do prolonged periods of monotonous, automated driving affect the driver state?
- a) Is there a difference in the progression of driver states due to different traffic densities during automated driving?

RQ3 How do potential driver state changes affect the take-over performance?

RQ4 How does the effect from driver state changes on take-over performance compare to the effect of different situations?
- a) How does the take-over performance compare to manual drivers experiencing identical situations?

Chapter 5

Driver State

Experiment 2: Effect of NDRTs

Pre-published in Radlmayr, Fischer, and Bengler (2019).

RQ2 How do different NDRTs affect the driver state?
- a) Is there a difference in the engagement in NDRTs depending on the instruction?

RQ3 How do potential driver state changes affect the take-over performance?

RQ4 How does the effect from driver state changes on take-over performance compare to the effect of different situations?

Chapter 6

Driver State

Modeling of take-over performance with mixed effect models

RQ5 How do effects from e.g. driver state changes, situational factors and NDRTs on take-over performance compare to the effect of predisposition and individual differences of drivers?

Chapter 7

Figure 3.2: Overview of Experiments 1 and 2 and the modeling approach focusing on driver state changes and their affect on take-over performance.

A quantitative measurement of driver state changes is essential to allow an empirically based insight into the underlying human factor problems. Based on the literature review, driver monitoring systems with delimitation to e.g. an expert rating of driver state changes are chosen for all four experiments. Eye-tracking provides the most inside into relevant driver state changes and was utilized in all experiments. Seat pressure mats are used in Experiments 1 and 2 as novel method of assessing driver state changes hypothesized to both detect changes in activity from prolonged durations of CAD as well as an engagement in NDRTs. The boxes provide information on the link between the general research questions RQ1-7 in this thesis and their specific differentiation in the experiments and chapters. The individual chapters on the four experiments differ in length and depth. Experiments 2 and 4 have been pre-published to this thesis. In comparison to a full depiction of the individual method, results and discussion for Experiments 1 and 3 in Chapters 5 and 9, Chapters 6 and 8 on Experiments 2 and 4 only provide a brief summary. For additional details on Experiment 2, refer to Radlmayr, Fischer, and Bengler (2019).

HMI

Experiment 3: Effect of peripheral monitoring during CAD

Pre-published in Radlmayr, Brüch, et al. (2018).

RQ6 How can the HMI for CAD and the take-over be optimized by enabling the possibility of peripheral monitoring while engaging in NDRTs during CAD?

a) Does the situation awareness of drivers differ during CAD depending on the possibility of peripheral monitoring?

b) Does the potential effect of situation awareness differences affect the take-over performance?

Chapter 8

HMI

Experiment 4: Optimization of the HMI for the take-over

RQ7 How can the HMI be optimized by offering additional information on the specific situation during the take-over process?

a) Does offering additional information in the HUD during the take-over affect the take-over performance?

b) How does the effect from additional information compare to the effect of different situations?

Chapter 9

Figure 3.3: Overview of Experiments 3 and 4 focusing on the optimization of the HMI for CAD and take-over process.

Experiment 4 is portrayed in detail in Radlmayr, Brüch, et al. (2018). The modeling approach in Chapter 7 is based on data from Experiments 1 and 2 and is also described in full detail in thesis.

The main contribution of this work consists of providing empirically based and critically discussed answers to the research questions, helping to fill the research gap on fundamental human factors of CAD.

4 General method

This chapter provides the general method that was kept identical throughout all four experiments. The experiments were all conducted in the static simulator of the Chair of Ergonomics at the Technical University of Munich, sharing key elements concerning the design of experiments. General principles of this work concerning the analysis of data and assessment of results are also provided.

Most of the younger participants were students at the TUM. The overall sample featured very few foreign participants, showing a homogeneous distribution regarding nationality with mainly German participants.

4.1 Experimental Setup

The setup in the static simulator was based on methodological assumptions discussed and developed in the project Ko-HAF (*Kooperatives, hochautomatisiertes Fahren – Ko-HAF*, 2018). Generic requirements for the HMI utilized for the take-over in CAD were based on the following features (*Kooperatives, hochautomatisiertes Fahren – Ko-HAF*, 2018):

- Transparent and continuous display of system state consisting of, at least, the states "Not available", "Available", "Active", and "Request to intervene".
- In case the system state consisted of an urgent, pressing and/or short-term display RtI, the display must be multi-modal to allow better perception.
- The multi modality should consist of e.g. a visual and acoustic signal, or voice output, while a haptic signal in the seat was considered optional.
- The display of the system states should allow highest priority for the RtI in case e.g. the display is used for featuring NDRTs.

Figures 4.1 and 4.2 depict the HMI including different system states and the RtI with a red steering wheel with two hands and a red text. This HMI was used as basis in all four

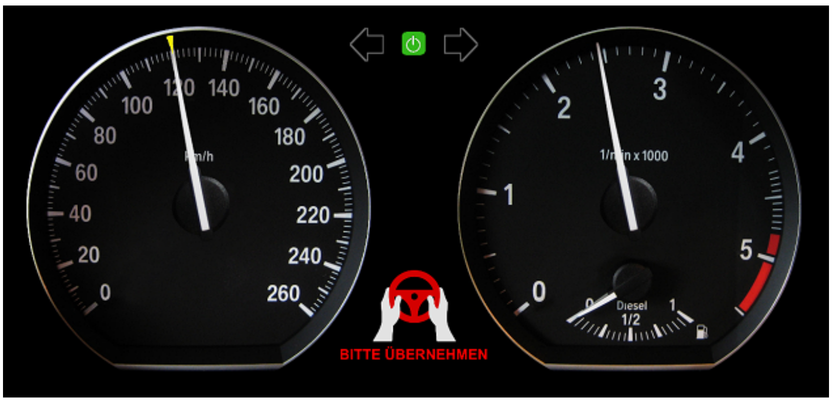


Figure 4.1: The generic HMI for system states including the RtI was derived from Ko-HAF (*Kooperatives, hochautomatisiertes Fahren – Ko-HAF*, 2018). The RtI is represented by a red steering wheel.

Figure 4.2: The icons show the different system states, "off", "available", and "active" of the generic HMI.

experiments while the experiments focusing on optimizing the HMI in Chapters 8 and 9 went beyond these generic requirements.

The take-over situations, their design and reasoning was based on the taxonomy from Gold et al. (2017). They can be distinguished in four categories, urgency, predictability, criticality and complexity of driver response. Each category can feature three levels, low, medium and high, shown in Table 4.1. Depending on the specific research focus, various combinations are recommended, as considered in Figure 4.2.

Table 4.1: Taxonomy for classifying and designing take-over situations for CAD (Gold et al., 2017). Depending on the specific research focus, different combinations of the categories were recommended.

Factors	Low (1)	Medium (2)	High (3)
Urgency	High time budget	Medium time budget	Small time budget
Predictability	Near-term detection of the system limit	Predictable, but occurrence dependent on situational conditions.	Known from backend, map, V2V-communication.
Criticality	Low safety risk	Increased safety risk	High safety risk
Driver Response	Low complexity (e.g. stabilizing)	Medium complexity (e.g. steering)	High complexity (e.g. lane change)

The design of the take-over situations in all four experiments relied on the taxonomy and adjusted the overall criticality accordingly. The situation "crash site" in Experiments 1 and 2 was chosen to focus on the maximal human performance in take-overs, while the situation "curve in heavy rain" in Experiment 4 targeted an analysis of results from optimizing the HMI.

Table 4.2: Recommendations for the combinations of the categories (Gold et al., 2017).

Research Focus	Urgency	Predictability	Criticality	Driver Response
Maximal Driver Performance	3	1	3	2-3
Mode Awareness / HMI	1-3	1-3	1-3	1-3
Driver Comfort	1-2	3	1	1-3

Figure 4.3: The figures show, from left to right, the simulator and the vehicle mockup, the three cameras of the eye-tracker SmartEye and the seat pressure mats and their implementation in the vehicle. Figures with friendly permission by Fabian Marco Fischer.

While other taxonomies for take-over situations have been proposed recently (e.g. McCall et al., 2018), these were not considered in this thesis.

The static simulator used for all experiments featured a BMW 6-series convertible with various projectors, allowing a 180 ° field of view horizontally, rear-view mirrors for drivers and a high immersion. The simulation was implemented with the software *SILAB*. The eye-tracking system consisted of a fixed-based, three camera-setup eye-tracker from *SmartEye*. Infrared flashes were utilized to better identify eye and facial features. The seat pressure mats from *XSensor* allowed a resolution of 12,7 mm and were used independent for both seat and backrest area. Figure 4.3 shows the mock-up from the outside and the interior with the eye-tracking system and the seat pressure mats.

4.2 Measures

The phrasing of take-over performance is taken from Gold (2016) and integrates both time aspects, such as reaction times and quality aspects, such as maximal accelerations or the TTC.

Some measures from previous research on take-over performance are not considered. The reaction time to the first glance away from a visual NDRT and subsequent reactions times not including TOT, e.g. time until the hands reach the steering wheel or the feet reach the pedals, will not be utilized in this thesis. These reaction times are likely to follow standard stimulus-response patterns after the RtI (Gold, 2016). The modeling of the "first-glance-away-from-a-NDRT-reaction time" showed mean values of .47 s with a standard deviation (SD) of .11 s and the intercept to be the best predictor (Gold, 2016). Consequently, focus is put on the TOT because it contains both the stimulus-reaction patterns and the cognitive process of re-entering the driver-vehicle control loop offering the most relevant insight concerning driver reactions during the take-over process.

The measures used for assessing state changes including eye-tracking, e.g. PERCLOS and blink frequency, and the measures to evaluate take-over performance, e.g. TOT, TTC and accelerations are based on a broad foundation of past research detailed in Chapter 2. Information on the individual combinations of measures for every experiment including measures that were disregarded due to e.g. technical problems or not meeting tracking quality thresholds, is found in the individual chapters.

Seat pressure mats

Seat pressure mats have typically been used for assessing seating comfort or discomfort, providing a variety of measures related to pressure, such as average pressure or peak pressure. In this thesis, the seat pressure mats were utilized as new way of assessing driver state changes. Two measures were identified to be most feasible in assessing potential state changes during automated driving. The contact area for both the seat and the backrest can be analyzed to quantify posture changes, especially regarding the upper torso. It was hypothesized that when drivers would e.g. engage in NDRTs located in the center console, visual attention in addition to their posture would change, potentially showing in the contact area in the backrest. Contrary, drivers relaxing during longer, monotonous periods of automated driving could show an increase in the contact area due to more contact with the seat needed to support a more relaxed posture. These effects in general are understood as motoric state changes following the framework in Figure 2.3.

The experiments 1 and 2 in Chapters 5 and Chapter 6 are aimed at quantifying the link between state changes measured using the seat pressure mats and take-over performance. The center of pressure (COP) is used to assess the activity of drivers both in the seat and the backrest. While the COP can be calculated for each sampling frame representing a spatial measure, changes in the COP represent drivers' activity. Murata, Koriyama, and Hayami (2012) provided a way of utilizing these changes of the COP. Equation 4.1 shows the equation used to calculate the COP for every sampling point where x_i is the x-coordinate at time x and x_{i-1} is the same coordinate one time step earlier, y respectively. The $\triangle$COP value represents the change in activity for drivers between every time step. Mean values were calculated for a period of one minute, and can be understood to represent the activity of drivers in the seat and backrest during that time. Changes in the driver state due to longer periods of automated driving can be evaluated by assessing the difference between e.g. the first and last minute of automated driving prior to a take-over.

$$\Delta COP = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \tag{4.1}$$

No absolute values but only the change between two periods is regarded to account for the general unrest individual drivers are showing during active CAD. The second minute (in contrast to the first) was considered to avoid data noise due to effects from settling into the seat right after activating the automated driving system.

Integrative measures of take-over performance – TOPS, TOC

The detailed assessment of take-over performance including various reaction times and quality measures motivated the introduction of new, more holistic metrics. Naujoks, Wiedemann, Schömig, Jarosch, and Gold (2018) suggested a rating of controllability based on the expert rating of videos of human and vehicle behavior during a take-over. While the assessment of controllability of take-overs must be considered concerning the future application of CAD, the rating was not considered in this thesis. Radlmayr, Ratter, et al. (2019) suggested the integration of different individual measures into three new parameters resulting in the take-over performance score (TOPS). The parameters representing the three most relevant assessments of take-overs are namely a vehicle guidance parameter, a mental processing parameter and a subjective rating parameter. The parameters allow the integration of a changing number of individual metrics, making it a more comprehensive evaluation of take-over performance. Further empirical validation of the TOPS is required before a substitution of the well-known individual measures seems feasible.

4.3 Results

Data handling, plotting and statistical analysis is carried out using the program *R* (R Core Team, 2018). For one- or two-sided statistical testing such as t-tests or an analysis of variance (ANOVA), the p-value is set to p=.05 throughout. Significant results are represented by a star (*) in the individual plot figures, where one star represents $p < .05$, two stars (**) represent $p < .01$ and three stars (***) represent $p < .001$.

A Shapiro-Wilks Test (SWT) is conducted to check if the sample represents a normally distributed population in the individual experiments. In case the SWT shows a significant result, Salkind (2010) show a high robustness of the ANOVA concerning a violation of the normal distribution for medium to large sample sizes (see also central limit theorem). This is the case for all the experiments in this thesis. Consequently, an ANOVA is carried out despite significant results from a SWT, in order to avoid losing insight of interaction effects. Results of the SWTs are reported regardless to allow an understanding of the magnitude of the effect.

Levene-tests are conducted in combination with a F-max test to evaluate the homogeneity of variance for the between factor. In case of significant results from the Levene-test, the p-value was adjusted in case the F-max-test showed a ratio larger than ten between the smallest and largest group variance and group size ratios are below four (Bühner & Ziegler, 2009, p. 518). This is not the case throughout this thesis.

Concerning sphericity, a Mauchly-test is applied and in case it showed significant results, the Greenhouse-Geisser-correction was used. Results of the ANOVA will indicate if the Greenhouse-Geisser-correction was used and are adjusted directly.

For the correction of pairwise comparisons following significant ANOVA results, the Holm-Bonferroni correction was used instead of the Bonferroni correction, since the Holm-Bonferroni correction is uniformly more powerful compared to the classical method (Holm, 1979; Singmann, 2017). Regarding significant results for the interaction effect vs. the main effects in mixed design ANOVAs, the procedure suggested by Bortz and Döring (2016, p. 714) was contemplated for the discussion in Chapters 5 and 9. In case the ANOVA yielded a significant result for the interaction of the factors, both main effects were considered for ordinal interactions, only one main effect was considered for a hybrid interaction and no main effects were considered for disordinal interactions. The type of interaction was determined by analyzing the estimated means in the respective plot figures in the appendix. Both interaction and main effect results were reported regardless of significance.

5 The effects of prolonged conditionally automated driving on driver state and take-over performance

The first experiment[1] was conducted to provide answers to the research questions 1, 3 and 4, see Figure 5.1. Chapter 2 conveyed a comprehensive overview of effects on take-over performance. Important effects such as traffic density or the time budget were already researched and allow a first understanding of situational factors. In addition, the existence of a growing number of data sets on take-overs motivated modeling efforts. Following the conclusions based on these models, evidence was revealed that these models benefit from considering data from driver monitoring (Gold, 2016). A large majority of empirical studies focused on the limit of maximal human performance as fallback in a take-over by e.g. amending external factors, such as situation criticality. NDRTs were part of the scope of the scientific studies and focused on the resulting take-over performance without including a thorough understanding of prior driver state changes. The more recent years have seen a large increase in studies focusing on driver state changes. Typically, studies focused either on NDRTs or the effects of drowsiness with this experiment focusing on drowsiness as consequence of prolonged CAD. The connection between prolonged driving both for manual and CAD and a progression of drowsiness was well researched in the last years.

Concerning the duration of CAD, a fixed-time design of experiments was implemented to reduce experimental variance. Results from research on manual driving showed a greater increase of drowsiness in more monotonous road environments compared to more frequented ones (Thiffault & Bergeron, 2003). It was hypothesized, that the progression of the driver state is depending on the traffic surroundings, more precisely the number of additional vehicles. Following the well-known results of traffic density on take-over performance (see Chapter 2.2.1), traffic density could also moderate other effects in take-over situations. Thus, driver state changes should be analyzed in take-over situations differing in their situational criticality.

In manual driving, a period of 30 minutes of driving would not see similar effects of duration since drivers would be in the loop. Concerning CAD and known effects from an onset of drowsiness (see Chapter 2.2.3), 30 minutes of automated driving without the possibility to engage in NDRTs was hypothesized to be sufficient to motivate a progression of drowsiness.

Concluding, Experiment 1 focused on the effect of prolonged CAD on the take-over performance in different situations. To allow a more comprehensive understanding of the driver state changes, the traffic density varied in this experiment. Results from manual drivers experiencing identical situations served as basis for the classification of absolute values concerning e.g. accelerations or TTCs. To allow an assessment of driver state changes, eye-tracking and seat pressure mats were used to quantify these changes during CAD while well-known metrics of take-over performance were used to asses reaction times and the quality of the take-over.

[1] The experiment was designed and conducted with the assistance of Lisa Scherer as part of her master's thesis (Scherer, 2016)

Driver State

Experiment 1: Effect of prolonged duration of CAD

RQ1 How do prolonged periods of monotonous, automated driving affect the driver state?
   a) Is there a difference in the progression of driver states due to different traffic densities during automated driving?

RQ3 How do potential driver state changes affect the take-over performance?

RQ4 How does the effect from driver state changes on take-over performance compare to the effect of different situations?
   a) How does the take-over performance compare to manual drivers experiencing identical situations?

Chapter 5

Figure 5.1: Research questions of Experiment 1.

5.1 Research questions

The main research questions were motivated by the lack of a detailed look on changing driver states during prolonged CAD in addition with understanding these effects in different situations. The literature overview portrayed in Chapter 2 served as reasoning narrowing down the research questions, see Figure 5.1. The derivation of null and alternative hypotheses was based on these research questions to allow the analysis of results using inferential statistics. Serving as example, the first alternative hypotheses is presented here. Analysis of results is based implicitly on underlying pairs of hypotheses throughout this chapter. Concerning the first research question, the alternative hypothesis is deduced:
H11: The driver state shows changes between 5 and 30 minutes of CAD.
Acceptance or rejection of the hypotheses was always based on the outcome of the corresponding statistical analysis. Based on the derivation of the exemplary H11, the research questions, 1, 3 and 4 were all translated into pairs of hypotheses but are not detailed here.

5.2 Method

Sample

A total of 60 participants took part in the experiment. Due to technical issues, the data of three participants had to be excluded from analysis. The remaining 57 participants (35 males and 22 females) were between 19 and 70 years old, with an average age of 32.8 years (SD: 13.2 years) and a median of 28 years. All participants held a regular driver's license with an average time of possession of 14.8 years (SD = 12.8 years). Fifteen participants reported to have experienced CAD in a simulator setup at least once.Twenty-one participants reported to have impaired vision, which was either corrected by glasses or contact lenses.

Experimental Setup

The experiment was conducted in the static driving simulator of the Chair of Ergonomics of the Technical University of Munich. The design of the experiment consisted of a mixed, incomplete setup which featured four independent variables. Table 5.1 shows the four factors, their levels and their integration into the experimental setup.

Table 5.1: Overview of the four independent variables and the way they were incorporated in the design of experiment.

Factor	Levels	Design
Level of Automation	**2** (manual driving, CAD)	Between subjects factor
Traffic Density	**2** (0 and 20 vehicles/km)	Between subjects factor
Type of situation	**3** (crash site, construction site, interstate crossing)	Within subjects factor
Duration of CAD	**2** (5 and 30 minutes)	Within subjects factor

The level of automation was split into two levels, manual driving and CAD. While the manual drivers represent the baseline concerning absolute values, there were a total of three groups: one with no automation and two featuring CAD. In addition to the between subjects factor Level of Automation, different traffic density conditions were introduced to the groups. While the manual drivers experienced 20 vehicles per kilometer, the two CAD-groups differed concerning their traffic density. One group featured 20 vehicles per kilometer randomly placed by the simulation software throughout the track, while the other CAD-group had no additional traffic throughout the simulation. This was implemented to analyze potential differences in participants' state during CAD. It was hypothesized, that 30 minutes of CAD without the possibility of engaging in NDRTs would lead to underload, potentially including an onset of drowsiness. The development of the driver state without a NDRT could be different, if drivers would use the possibility of monitoring surrounding traffic in the 20 vehicles/km-group. The analysis in the result section consequently split the group comparison accordingly. For the effect of Level of Automation, the manual group was only compared to the CAD group with surrounding traffic (20 vehicles/km). For the effect of Traffic Density, only the two CAD groups were compared.

The two within subjects factors Type of Situation and Duration of CAD were interlinked and can be understood by considering the procedure of the experiment. After an interval of CAD (5 or 30 minutes long), the take-over takes place in a specific situation (crash site or construction site or interstate crossing). Consequently, by having participants experience each of the three situations once, the duration of CAD prior to each take-over was defined to be 5 minutes twice and 30 minutes once. This kept the duration of the overall experiment to a reasonable interval and allowed a group comparison between take-overs after 5 and 30 minutes. The sequence of both the take-over situations and the duration of CAD was permuted to avoid sequence effects. Table 5.2 shows the combination of the two between subjects factors into the incomplete group setup. In addition to the between subjects factors, the design of experiment included two within subjects factors, Type of Take-over Situation and Duration of CAD. Participants experienced every situation only once, to account for a total of three take-overs per participant. Every take-over was preceded by a period of CAD, either being 5 or 30 minutes long. Figure 5.2 shows the incomplete design of experiments including all four independent variables.

Prior to every experimental drive, participants had the chance to familiarize themselves with the simulator dynamics including an exemplary take-over. The track was setup to be a stretch of German interstate with three lanes, with the familiarization drive lasting

Table 5.2: The combination of the two independent variables "automation level" and "traffic density" and their aggregation into the three groups.

Groups	Automation level	Traffic density
CAD0	CAD	0 vehicles/km
CAD20	CAD	20 vehicles/km
Manual20	Manual	20 vehicles/km

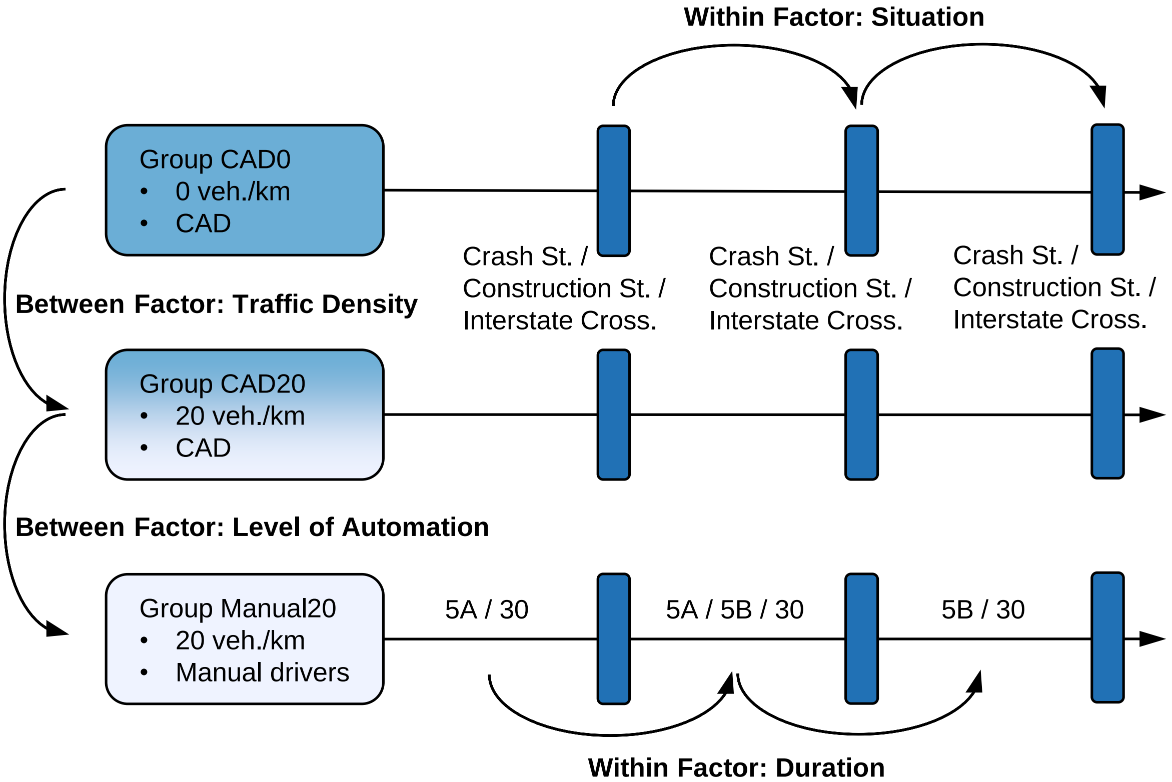


Figure 5.2: Overview of the design of experiments, including all independent variables and their interaction. The analysis of results including several ANOVAs for each metric is a result of the possible combinations of between and within subjects factors.

approximately 10 kilometers. The manual group with no CAD available were instructed to follow a target lane specified throughout the experiment by the operator via intercom to ensure that participants in this group were experiencing the "take-over" situations identical to the groups with CAD. A RtI was presented in the manual group as well and was explained as additional warning assistance in dangerous situations for manual drivers. The experimental track was approximately 83 kilometers long and lasted 42 minutes. The maximal automation speed was set to 120 km/h. The automation was programmed to execute lane changes by itself. Only very few were executed in the CAD group without additional traffic and more were executed according to the traffic situation in the group with additional traffic. Both CAD groups featured oncoming traffic on the other side of the median strip. All take-over situations featured a time budget of seven seconds, translating into 233 meters of distance from the moment the RtI was issued to the system limit. The RtI was part of the generic HMI, as displayed in the general method in Chapter 4. Perception of the system limits prior to the RtI was not possible due to the track layout.

Figure 5.3: The three take-over situations, the construction site, the interstate crossing and the crash site (left to right). The figures show the take-over situation only few meters prior to the system limit.

The display was always active and showed the status of the automation. Following the classification of Gold et al. (2017), the three situations were designed to differ in their overall criticality. The crash site featured the highest criticality, the interstate crossing was set to represent a medium overall criticality and the construction site was the least critical take-over situation. In the interstate crossing, participants were instructed to conduct a lane change via the intercom using a computer-generated voice starting at the time of the RtI. The auditory message lasted two seconds, technically shortening the available time budget of the situation. Considering the literature and typical TOTs (Gold, 2016), the auditory message was classified to not be relevant concerning reaction times. An auditory message is not interfering with the main modality of manual driving (visual) in addition to being perceived in the very beginning of the take-over process, typically before a conscious reaction would be issued by drivers (mean TOTs (Gold, 2016)). Figure 5.3 provides an overview of the three take-over situations right before the system limit is met.

Measures

The static simulator and the software SILAB allowed the acquisition of all relevant vehicle dynamics in 120 Hz. This includes a fully simulated dynamics model of the ego vehicle, for example providing accelerations and chassis movement. Eye-tracking was used to evaluate the driver state throughout the experiment. Prior to analysis, a data error was discovered for the measurement of the eye-lid opening value due to a casting error while sending the data to the combined data storage. The original data was recovered by conducting several new measurements of eye-lid opening data and comparing them to the correct values prior and after sending them. Throughout the scale of values of eye-lid opening, the correct values were retracted.

Seat pressure mats were also implemented in the vehicle mockup, allowing the measurement of pressures in seat and backrest. Video cameras were used to record the interior of the vehicle to allow a comprehensive view on participants, since the eye-tracker was focused on the head area. Figure 4.3 shows the simulator setup including eye-tracker and seat pressure mats. Table 5.3 gives an overview of the measures used in this work. The list differentiates between measures for assessing the driver state during CAD prior to a take-over and the measures for take-over performance. The assessment of take-over performance is limited to the interval from the RtI until drivers had solved the situation by passing the system limit.

Table 5.3: Summary of dependent variables used for assessing the driver state and the take-over performance.

Driver state during CAD	Take-over performance
Eye-Tracking	**Time aspects**
• Standard deviation of the horizontal gaze position (horizontal gaze dispersion, HGD) (per minute) [m] • Percentage eyes on road (PEOR) (per minute) [%] • Percentage of eyes closed (PERCLOS) (per minute) [%] • Average blink duration (per minute) [s] • Average blink frequency (per minute) [Hz]	• Minimal time to the first deliberate action / take-over time (steering wheel >2 ° or brake pedal >10 % or deactivation of automation by button) (TOT) [s]
Seat pressure mats	**Quality aspects**
• Changes in the center of pressure (COP) between second and last minute of CAD [%] • Changes in the contact area between second and last minute of CAD [%]	• Number of crashes [n] • Minimal (negative) longitudinal acceleration [m/s^2] • Maximal lateral acceleration (absolute value) [m/s^2] • Minimal time to collision (TTC) [s]
	Subjective ratings from participants
	• Criticality of situation [ ] • Complexity of situation [ ] • Urgency of situation [ ] • Comfort of the take-over (only CAD groups) [ ]

The SD of the horizontal gaze position represents the tracking activity of drivers in the horizontal direction. The SD increases, if drivers increase the frequency of looking left and right. This measure is also known as the horizontal gaze dispersion (HGD) and represents the visual tracking activity. The groups differed in their traffic density in addition with intervals of CAD lasting unevenly between 5 or 30 minutes without the possibility to engage in NDRTs. The HGD was analyzed to assess the visual tracking behavior of drivers during CAD.

For the calculation of the percentage of closed eyes during a one minute time interval (PERCLOS), the maximal value of the eye-lid opening was set to the maximal value during the first minute of CAD at the beginning of the experimental drive. Following the definition from Wierwille and Ellsworth (1994), the PERCLOS represents the number of frames within one minute were the eye-lids were between 80 %-100 % closed, with regard to the obtained maximal value of the eye-lid opening for each participant. With reference to the literature review on measures that capture an onset or progression of drowsiness (Chapter 2.3), PERCLOS can be understood to be the most prominent measure for drowsiness.

In addition to the PERCLOS, the blink duration can also be analyzed to assess an onset or progression of drowsiness caused by a prolonged duration of CAD. Corresponding to an increase in blink duration associated with an onset of drowsiness, the blink frequency decreases respectively for drowsy drivers (Dinges & Mallis, 1998).

Due to technical issues with the eye-tracking system, data were disregarded for analysis if the recording froze longer than one second without recovering afterwards. This was the case for three participants.

Procedure

Participants took approximately 75 minutes to pass the experiment. They were greeted at the simulator and started with a thorough introduction of the hardware, the sensors and the procedure after which participants signed a consent form. They filled out a demographic questionnaire after which they received a group-specific additional introduction depending if they would experience manual driving or CAD. In the following familiarization drive, participants could experience the simulator itself, vehicle dynamics and exemplary take-overs in the CAD-groups. Drivers in the manual group could experience the operator instructing the target lane in manual driving. They would also experience the take-over situation including the RtI which was explained as additional emergency assistance system prewarning drivers of critical situations. Before the experimental drive started, the eye-tracker and the seat pressure mats were calibrated. The experimental drive started and ended at a rest area. Participants were asked by the operator to activate the automation right after leaving the rest area and after every take-over after passing the system limit and answering the subjective rating questions via the intercom. After the experiment, participation was rewarded with 30 €.

5.3 Results

The analysis of data was conducted following the process described in Chapter 4.3. Since the experimental setup consisted of an incomplete design, the analysis of the take-over performance metrics, subjective ratings and driver state metrics could not be conducted using only one statistical test for each dependent variable but had to be conducted step-wise.

Concerning the two between factors, Traffic Density and Level of Automation, two individual ANOVAs were conducted. The level of automation was analyzed comparing the group Manual20 with the group CAD20, since both groups featured a traffic density of 20 vehicles/km. The group CAD0 did not experience any traffic during the experimental drive and could not be compared to the manual drivers. For the effect of traffic density, both CAD-groups were compared since they did not differ in their level of automation but featured different traffic densities.

In addition, the setup featured two within factors, Situation and Duration. The factor Duration technically consisted of two levels, differentiating between 5 and 30 minutes of

CAD. A duration of 5 minutes after 30 minutes of CAD greatly differs from 5 minutes right at the beginning of the experiment concerning potential effects of prolonged CAD. The duration of either 5 or 30 minutes was therefore differentiated in three levels, 5A, 5B and 30 minutes. The 5-minute duration 5B did not necessarily happen after a 30 minute duration but was labeled as the second time participants experienced 5 minutes of CAD. A potential effect of prolonged CAD could also manifest in the total time of CAD, either being 5 or 30 minutes before the first take-over, 10 or 35 minutes before the second take-over and 40 minutes before the third take-over. To account for the potential effect of total time of CAD, the analysis for the CAD groups featured an additional, surrogate within subjects factor, the trial number. The two factors Duration and Situation were not permuted completely since participants would have to experience a total of nine take-overs but experienced a total of three. Therefore, the analysis of within factors consisted of three mixed-design ANOVAs, one for the factor situation, one for duration of CAD and one for trial. Combined with the split between the two between factors, this resulted in a total of five mixed-design ANOVAs for each dependent variable. Results are depicted in tables to allow a good overview of corresponding results. In case specific metrics did not allow an analysis of e.g. level of automation, the corresponding ANOVAs were not calculated. This is the case for all metrics focusing on effects connected to CAD without the additional comparison to manual drivers and is indicated in the respective sections. Analysis of the factor situation is unreasonable for the driver state metrics, since the analysis focused on time intervals prior to the respective situations. For some metrics, the analysis of the surrogate factor trial was deemed unreasonable and is also indicated. Figure 5.4 provides an overview of all potential 5 ANOVAs per metric and the reasoning behind the step-wise analysis. A full analysis of potential interactions between all four independent variables was deemed less relevant for this individual experiment, but can be found in the modeling approach in Chapter 7.

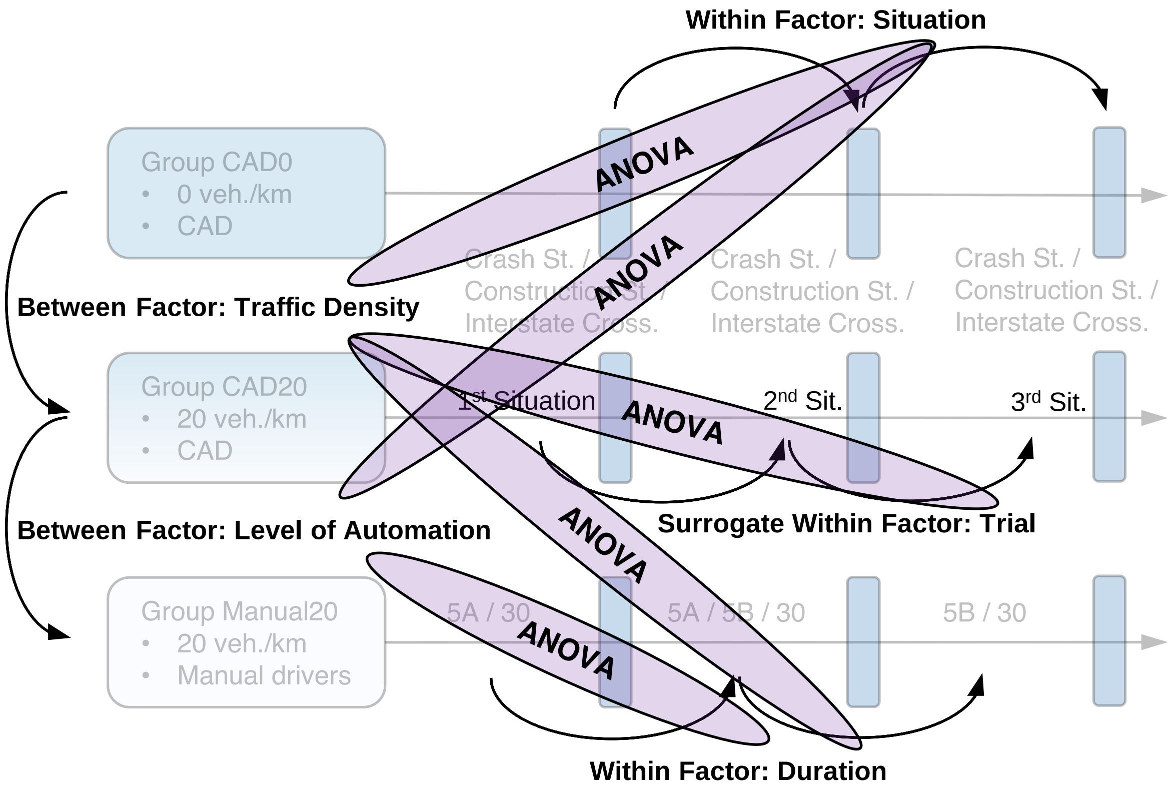


Figure 5.4: Overview of the step-wise analysis of metrics, including all potential five mixed-design ANOVAs, regarding both one between and one within factor. In case an ANOVA is disregarded for specific metrics, it is indicated in the respective section.

5.3.1 Measures of take-over performance

Crash rate

Two groups featured surrounding traffic in the take-over situations in addition with the crash site representing a system limit that required a full stop and/or a lane change in the take-over to avoid an accident involving the ego vehicle. Analysis of the crash rate can only be assessed for 54 participants, since the video data of three participants were not recorded due to technical issues. Out of a total of 162 take-overs including the group with manual drivers, six crashes were recorded. For the crash site situation only one accident into the left guard railing was recorded while one additional participant stopped in the middle of the left and middle lane, potentially presenting a hazardous obstruction for following vehicles. One participant in the interstate crossing collided into a vehicle on the right lane. In construction site, four participants collided with a simulated vehicle executing a lane change maneuver from the right lane to the middle lane, while an additional four participants executed a lane change maneuver to the left lane to avoid colliding into the simulated vehicle cutting into the ego lane. Due to the low numbers of total crashes, no statistical analysis was conducted.

Take-over time

The TOT was not analyzed for the manual drivers since they did not take-over, resulting in a total of three ANOVAs for TOT. The individual group means and SDs can be found in Table 5.4. Table A.1 shows the results from the SWTs, the Levene-tests and the F_{max}-

Table 5.4: Overview of the group and situation means and (SDs) in seconds for the TOT for the different situations and durations.

Within factor	CAD0	CAD20	Manual20
		TOT [s]	
Crash site	M = 1.93 (.52)	M = 1.83 (.73)	-
Construction site	M = 2.60 (1.60)	M = 2.18 (1.16)	-
Interstate Crossing	M = 1.40 (.64)	M = 1.91 (.80)	-
5A	M = 1.76 (.97)	M = 2.00 (1.05)	-
5B	M = 1.95 (1.20)	M = 1.99 (.97)	-
30	M = 2.23 (1.23)	M = 1.93 (.74)	-

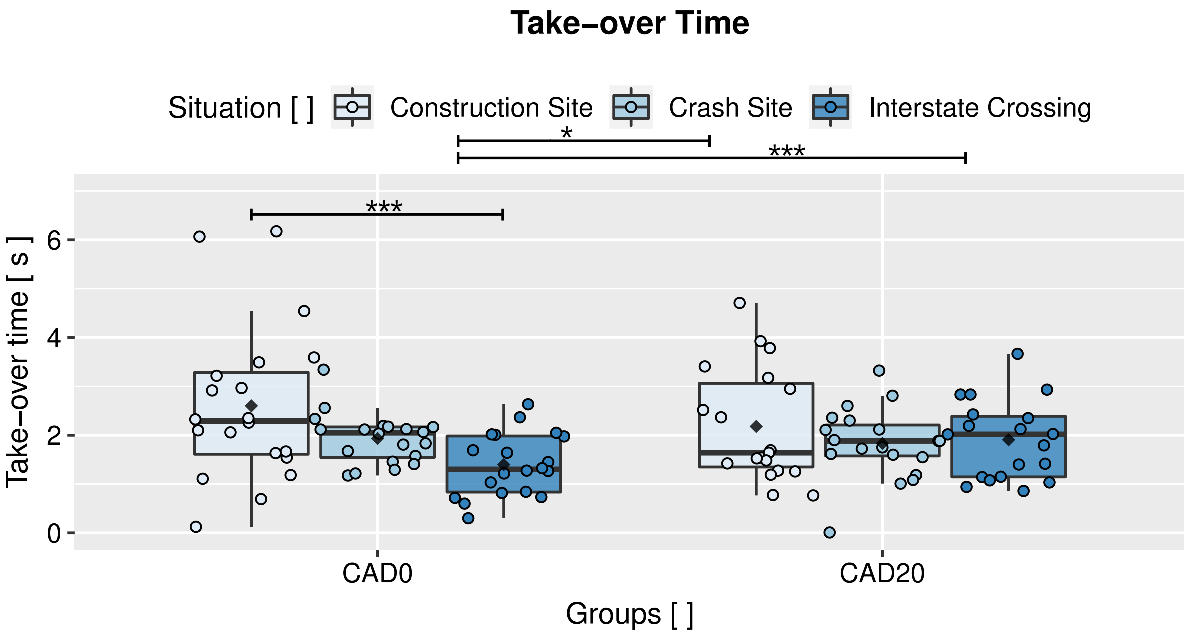


Figure 5.5: Plot of the TOT. Manual drivers were not analyzed since they did not take-over. n_{CAD0} = 60, n_{CAD20} = 57.

tests. Significant results were found for the factor situation and the interaction between traffic density and situation (Table 5.5). The data are plotted in Figure 5.5. Pairwise comparisons (Holm-Bonferroni corrected) revealed that the situation Construction Site differed significantly from the Crash Site ($p = .01$) and the Interstate Crossing ($p < .001$) by showing the highest TOTs. The pairwise comparisons (Holm-Bonferroni corrected) for the interaction revealed significantly higher TOTs ($p < .001$) in the Construction Site compared to the Interstate Crossing in case there is no additional traffic. The estimated means were plotted in Figure A.1 and underlined the results. No other significant results were discovered.

Time to collision

The TTC was analyzed as a measure of take-over quality. For the crash site, a collision with the vehicles blocking the lane was possible and the TTC represented the shortest

Table 5.5: Results from the ANOVAs conducted for TOT. If sphericity was violated, a Greenhouse-Geisser-correction was implemented and is indicated by (GG) succeeding the test statistic.

Factor combination	Main effect 1 (between)	Main effect 2 (within)	Interaction 1 x 2
		Take-over time	
Traffic density (1) and situation (2)	$F(1, 37) = .00$, $p = .98, \eta^2 < .01$	$F(1.68, 62.18) =$ $= 8.58$, (GG), $p = .001, \eta^2 = .09$	$F(1.68, 62.18) =$ $= 3.35$, (GG) $p = .05, \eta^2 = .04$
Traffic density (1) and duration (2)	$F(1, 36) = .08$, $p = .78, \eta^2 < .01$	$F(1.96, 70.56) =$ $= .81$, (GG), $p = .45, \eta^2 = .01$	$F(1.96, 70.56) =$ $= 1.27$, (GG) $p = .29, \eta^2 = .02$
Traffic density (1) and trial (2)	$F(1, 37) = .00$, $p = .98, \eta^2 < .01$	$F(1.93, 71.37) =$ $= .01$, (GG), $p = .99, \eta^2 < .01$	$F(1.93, 71.37) =$ $= 1.14$, (GG) $p = .32, \eta^2 = .02$

distance towards the obstacle right before the ego vehicle would steer clear of a potential collision divided by the current velocity. Technically, analysis of the TTC was restricted to the crash site since the construction site and the interstate crossing did not feature obstacles. In case participants would not take-over, in the construction site they would only gradually drift towards the side of the track and in the interstate crossing they would miss the appropriate exit. In order to compare situation, duration and group differences based on a more comprehensive data set, the TTC was also calculated for the construction site and the interstate crossing. In the construction site, the TTC represented the distance at the point where participants had taken over by deliberately steering, braking/accelerating or pushing the button to deactivate the automation divided by their current vehicle velocity. The point in time was identical to the time-budget minus the TOT. Consequently, the TTC was not analyzed for manual drivers in the construction site since no take-over took place. For the interstate crossing participants had to execute a lane change maneuver to take the appropriate exit. The distance from the moment participants steered clear of their original lane divided by their current velocity was utilized to calculate the minimal TTC in the interstate crossing. The calculation of TTC-values for the construction site (CAD groups) and the interstate crossing (all groups) was deemed reasonable in sight of the modeling approach depicted in Chapter 7. The TTC-values always represent the minimal TTC. Smaller values of the TTC represents a more critical take-over quality. Results from the Shapiro-Wilk-, Levene- and F_{max}-tests can be found in Table A.4. All five ANOVAs were conducted for the TTC. For the ANOVA on the automation level as between factor and the situations as within factor, only the crash site and the interstate crossing were considered for analysis. The group means and SDs can be found in Table 5.6. Key findings are listed to allow a more comprehensive understanding of accumulated results from the five ANOVAs (Table 5.7).

- Main effects
  - The first ANOVA considered the automation level as between factor and situation as within factor, in this case consisting only of the crash site and interstate

Table 5.6: Overview of the group and situation means (SDs) in seconds for the TTC for the different situations and durations of CAD. Manual drivers in the construction site were not regarded since no surrogate TTC could be calculated.

Within factor	CAD0	CAD20	Manual20
	Minimal TTC [s]		
Crash site	M = 2.27 (.87)	M = 1.29 (.49)	M = 1.68 (.53)
Construction site	M = 2.58 (2.20)	M = 3.97 (1.13)	-
Interstate crossing	M = 3.87 (1.16)	M = 2.10 (.98)	M = 3.63 (1.46)
5A	M = 2.53 (1.50)	M = 2.64 (1.22)	M = 2.80 (1.59)
5B	M = 3.05 (1.67)	M = 2.15 (1.55)	M = 3.05 (1.69)
30	M = 3.17 (1.79)	M = 2.62 (1.59)	M = 2.12 (.96)

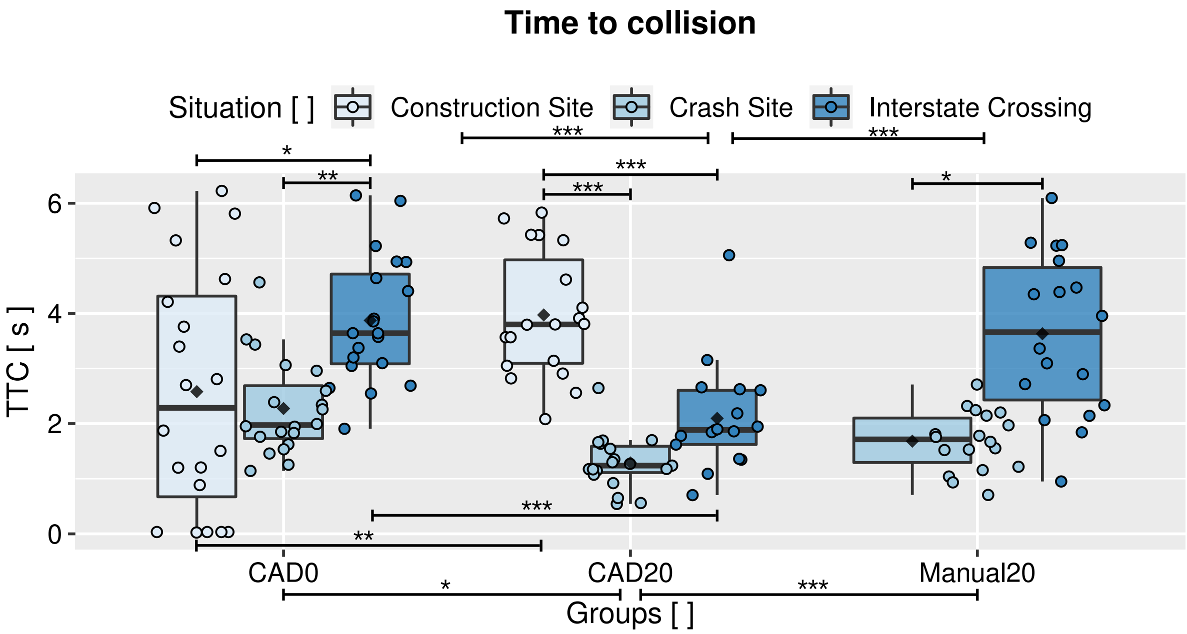


Figure 5.6: Plot of the minimal TTC between the RtI and the system limit. n_{CAD0} = 60, n_{CAD20} = 55, $n_{Manual20}$ = 36.

crossing. Results showed highly significant results for both main factors and significant results for their interaction.

  - Manual drivers showed significantly larger TTCs compared to participants taking over in the CAD20-group.
  - The second ANOVA of the factors Traffic Density and Situation considered all three situations. The factor Situation and the interaction between Traffic Density and Situation showed highly significant results. Pairwise comparisons (Holm-Bonferroni-corrected) for the main effect Situation revealed the crash site to show significantly smaller TTCs compared to the construction site ($p < .001$) and the interstate crossing ($p < .001$).
  - The main effect of Traffic Density either showed a tendency for significant effects or significant results for the main effect with the CAD20 group showing smaller TTCs compared to the CAD0 group.

- Interaction effects
  - The interaction between Automation Level and Situation showed the situation differences especially in the manual group, as well as in the CAD0 group and the CAD20 group.
  - Concerning the interaction effect between Traffic Density and Situation, pairwise comparisons (Holm-Bonferroni-corrected) were also conducted as follow-up tests. The between factor Traffic Density showed significant differences for the situation construction site ($p < .01$) and for the interstate crossing ($p < .001$). Contrary to the construction site, where the group without traffic (CAD0) revealed lower TTCs, the same group, CAD0, revealed higher TTCs compared to the group CAD20 in the interstate crossing.
  - Concerning situational differences, in the CAD0-group, the interstate crossing showed significantly higher TTCs compared to the construction site ($p = .02$) and the crash site ($p < .01$).
  - In the CAD20-group the construction site showed significantly higher times to collision compared to the crash site ($p < .001$) and the interstate crossing ($p < .001$).

Plots of the estimated means for the TTC clarifying the significant results can be found in Figure A.5. The ANOVA for automation level and duration led to an unbalanced design for the mixed ANOVA and was not analyzed. For the effect of prolonged CAD, only the ANOVA for Traffic Density and Duration was considered (Table 5.7). Results underlined significant differences for the factor Traffic Density, but not for the duration and the interaction. Results for Traffic Density and Trial (Table 5.7) showed a tendency for Traffic Density, but no significant results for Trial and the interaction between them.

Longitudinal acceleration

Concerning also the quality of the take-over, the longitudinal acceleration was analyzed. Values represent the maximal negative accelerations during the take-over. This can also be understood to represent the maximal brake acceleration. Following the F_{max}-test in Table A.2, the p-value was not adjusted. For the longitudinal acceleration, all possible five ANOVAs were conducted. The group means and SDs can be found in Table 5.8 and the

Table 5.7: Results from the ANOVAs conducted for the TTC. If sphericity was violated, a Greenhouse-Geisser-correction was implemented and is indicated by (GG) succeeding the test statistic.

Factor combination	Main effect 1	Main effect 2	Interaction 1 x 2
		Min. time to collision	
Automation level (1) and situation (2)	$F(1,33) = 20.35,$ $p < .001, \eta^2 = .21$	$F(1,33) = 31.90$ $p < .001, \eta^2 = .35$	$F(1,33) = 5.55$ $p = .02, \eta^2 = .09$
Traffic density (1) and situation (2)	$F(1,35) = 3.42$ $p = .07, \eta^2 = .03$	$F(1.69, 58.99) =$ $= 13.81,$ (GG), $p < .001, \eta^2 = .22$	$F(1.69, 58.99) =$ $= 15.24,$ (GG), $p < .001, \eta^2 = .23$
Automation level (1) and duration (2)	-	-	-
Traffic density (1) and duration (2)	$F(1,34) = 4.53$ $p = .04, \eta^2 = .02$	$F(1.99, 67.80) =$ $= .54,$ (GG), $p = .59, \eta^2 = .01$	$F(1.99, 67.80) =$ $= .66,$ (GG), $p = .52, \eta^2 = .02$
Traffic density (1) and trial (2)	$F(1,35) = 3.42$ $p = .07, \eta^2 = .02$	$F(1.97, 68.88) =$ $= 1.84,$ (GG), $p = .17, \eta^2 = .04$	$F(1.97, 68.88) =$ $= .17,$ (GG), $p = .84, \eta^2 < .01$

Table 5.8: Overview of the group and situation means and (SDs) for the longitudinal acceleration in meters per seconds squared for the different situations and durations of CAD.

Within factor	CAD0	CAD20	Manual20
	Longitudinal acceleration [m/s^2]		
Crash site	M = -3.85 (3.40)	M = -3.87 (3.59)	M = -4.68 (3.82)
Construction site	M = -3.03 (3.17)	M = -6.47 (2.21)	M = -4.94 (3.06)
Interstate crossing	M = -6.46 (1.72)	M = -6.48 (2.32)	M = -7.15 (1.90)
5A	M = -4.01 (3.25)	M = -5.72 (3.02)	M = -5.53 (3.01)
5B	M = -4.09 (3.16)	M = -5.32 (3.20)	M = -6.20 (2.88)
30	M = -5.24 (3.12)	M = -5.77 (2.90)	M = -5.09 (3.66)

Table 5.9: Results from the ANOVAs conducted for the longitudinal acceleration. If sphericity was violated, a Greenhouse-Geisser-correction was implemented and is indicated by (GG) succeeding the test statistic.

Factor combination	Main effect 1 (between)	Main effect 2 (within)	Interaction 1 x 2
	Longitudinal acceleration		
Automation level (1) and situation (2)	$F(1,34) = .00$, $p = .96, \eta^2 < .01$	$F(1.67, 56.87) =$ $= 9.59$, (GG), $p < .001, \eta^2 = .11$	$F(1.67, 56.87) =$ $= 3.19$, (GG), $p = .06, \eta^2 = .04$
Traffic density (1) and situation (2)	$F(1,37) = 3.36$, $p = .07, \eta^2 = .04$	$F(1.80, 66.55) =$ $= 11.26$, (GG), $p < .001, \eta^2 = .13$	$F(1.80, 66.55) =$ $= 6.26$, (GG), $p < .01, \eta^2 = .08$
Automation level (1) and duration (2)	$F(1,34) = .00$, $p = .96, \eta^2 < .01$	$F(1.88, 64.00) =$ $= .07$, (GG), $p = .92, \eta^2 < .01$	$F(1.88, 64.00) =$ $= .67$, (GG), $p = .51, \eta^2 = .01$
Traffic density (1) and duration (2)	$F(1,36) = 2.69$, $p = .11, \eta^2 = .04$	$F(1.93, 69.49) =$ $= 1.05$, (GG), $p = .35, \eta^2 = .02$	$F(1.93, 69.49) =$ $= .53$, (GG), $p = .59, \eta^2 < .01$
Traffic density (1) and trial (2)	$F(1,37) = 3.36$, $p = .07, \eta^2 = .04$	$F(1.92, 71.06) =$ $= 1.95$, (GG), $p = .13, \eta^2 = .03$	$F(1.92, 71.06) =$ $= .02$, (GG), $p = .97, \eta^2 < .01$

data including significant findings are visualized in Figure 5.7. The following list provides an overview on most important findings from all the ANOVAs. Details on the test statistics can be found in Table 5.9.

- Main effects
  - The factor Situation showed highly significant results with a tendency for the interaction. Pairwise comparisons (Holm-Bonferroni-corrected) for the factor situation revealed a significant difference between all three situations (crash vs. construction p = .04, crash vs. crossing p < .001, construction vs. crossing p = .05). The interstate crossing showed the most severe brake accelerations, followed by the construction site and the crash site.
  - The factor Traffic Density showed a tendency for significant results, with the group CAD20 showing a tendency towards more severe brake reactions.
- Interaction effects
  - Pairwise comparisons (Holm-Bonferroni-corrected) were conducted for the tendency of a significant interaction effect, showing a significant difference between the situation crash site and the construction site (p = .01) and the crash site and the interstate crossing (p = .01). Participants braked significantly less in the crash site compared to the other situations in the group CAD20.
  - The interaction between traffic density and situation showed significant results.

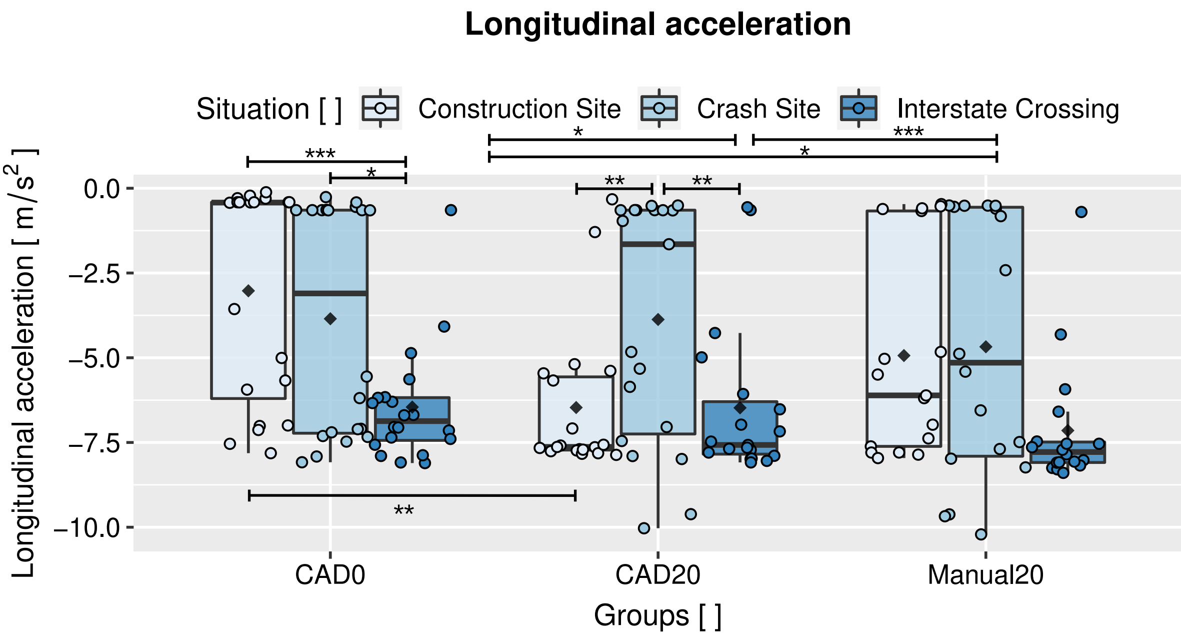


Figure 5.7: Plot of the most negative values of the longitudinal acceleration between the RtI and the system limit representing the maximal brake acceleration. n_{CAD0} = 60, n_{CAD20} = 57, $n_{Manual20}$ = 53.

- The between factor Traffic Density showed significant differences ($p < .01$) only in the construction site with the CAD20 group showing more severe brake reactions.
- In the CAD0-group, the interstate crossing showed significantly more severe brake reactions compared to the crash site ($p = .02$) and the construction site ($p < .001$).
- In the CAD20-group the crash site showed significantly less severe brake accelerations compared to the construction site ($p = .02$) and the interstate crossing ($p < .01$).

Figure A.2 shows the main and interaction effects for the situations following the results from Table 5.9. The overall effect of time was analyzed with an additional mixed-design with the factor trial number as within factor. Results did not show any additional significant effects (Table 5.9). The second set of ANOVAs analyzed the effect of prolonged CAD in combination with the level of automation and the traffic density according to the analysis of situation differences. No significant results were found for Level of Automation, Duration and their interaction. The ANOVA considering the factors Traffic Density and Duration also revealed no significant results for both main effects and their interaction.

Lateral acceleration

In addition to the longitudinal acceleration, the maximal lateral acceleration between RtI and system limit was analyzed as measure of take-over quality. In case participants stabilized the vehicle in lane or conducted a lane change, higher values of the lateral acceleration represent more dynamic maneuvers. Following the significant Levene-test and the follow-up F_{max}-test in Table A.3, the p-value was not adjusted. According to the

Table 5.10: Overview of the group and situation means (SDs) for the lateral acceleration in meters per seconds squared for the different situations and durations of CAD.

Within factor	CAD0	CAD20	Manual20
		Max. lateral acceleration [m/s^2]	
Crash site	M = 3.55 (.82)	M = 4.67 (1.42)	M = 4.06 (1.51)
Construction site	M = 1.52 (1.47)	M = 1.57 (1.68)	M = 1.90 (1.53)
Interstate crossing	M = 1.73 (.77)	M = 2.65 (1.84)	M = 1.82 (.72)
5A	M = 2.21 (1.27)	M = 2.97 (2.00)	M = 2.29 (1.56)
5B	M = 2.53 (1.54)	M = 2.87 (1.99)	M = 2.65 (1.41)
30	M = 2.08 (1.42)	M = 3.05 (2.34)	M = 2.87 (1.99)

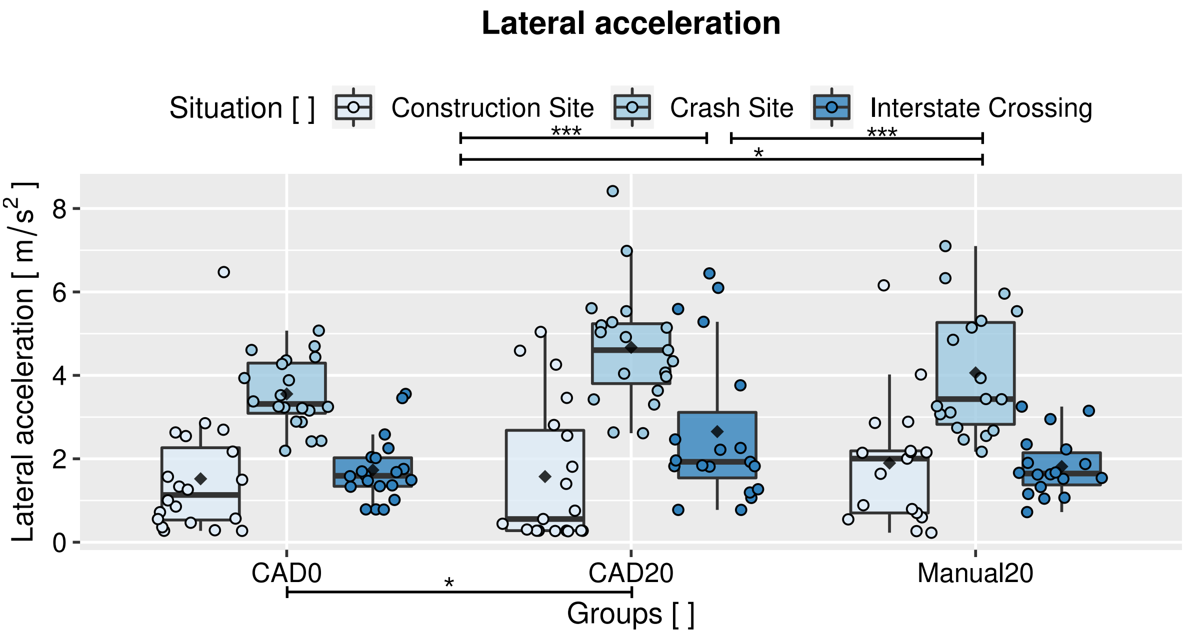


Figure 5.8: Plot of the maximal lateral acceleration between the RtI and the system limit. $n_{CAD0} = 60$, $n_{CAD20} = 57$, $n_{Manual20} = 53$.

analysis of the longitudinal acceleration, all five ANOVAs were also conducted for the lateral acceleration. The group means and SDs can be found in Table 5.10. Significant results were found for the main effects traffic density and situation. No significant interaction results were revealed throughout all ANOVAs. The key findings are listed in the following while details on the test statistics can be found in Table 5.11.

- The ANOVA with the factors Automation Level and Situation revealed a significant result for the factor Situation. Pairwise comparisons (Holm-Bonferroni-corrected) for the factor Situation revealed a significant difference between the crash site and the two other situation (crash vs. construction $p < .001$, crash vs. crossing $p < .001$). The crash site showed significantly higher lateral accelerations compared to the other situations (see Figure 5.8 and the corresponding plot of estimated means in Figures A.3 and A.4).

Table 5.11: Results from the ANOVAs conducted for the lateral acceleration. If sphericity was violated, a Greenhouse-Geisser-correction was implemented and is indicated by (GG) succeeding the test statistic.

Factor combination	Main effect 1 (between)	Main effect 2 (within)	Interaction 1 x 2
	Max. lateral acceleration		
Automation level (1) and situation (2)	$F(1,34) = 1.54$, $p = .22, \eta^2 = .02$	$F(1.88, 63.96) =$ $= 37.99$, (GG), $p < .001, \eta^2 = .37$	$F(1.88, 63.96) =$ $= 2.21$, (GG), $p = .12, \eta^2 = .03$
Traffic density (1) and situation (2)	$F(1,37) = 5.66$, $p = .02, \eta^2 = .06$	$F(1.96, 72.65) =$ $= 42.85$, (GG), $p < .001, \eta^2 = .39$	$F(1.96, 72.65) =$ $= 1.93$, (GG), $p = .15, \eta^2 = .03$
Automation level (1) and duration (2)	$F(1,34) = 1.54$, $p = .22, \eta^2 = .02$	$F(1.82, 62.00) =$ $= .08$, (GG), $p = .91, \eta^2 < .01$	$F(1.82, 62.00) =$ $= .10$, (GG), $p = .89, \eta^2 < .01$
Traffic density (1) and duration (2)	$F(1,36) = 5.26$, $p = .03, \eta^2 = .04$	$F(1.75, 62.94) =$ $= .08$, (GG), $p = .90, \eta^2 < .01$	$F(1.75, 62.94) =$ $= .32$, (GG), $p = .70, \eta^2 < .01$
Traffic density (1) and trial (2)	$F(1,37) = 5.66$, $p = .02, \eta^2 = .04$	$F(1.88, 69.54) =$ $= .18$, (GG), $p = .82, \eta^2 < .01$	$F(1.88, 69.54) =$ $= 1.16$, (GG), $p = .32, \eta^2 = .02$

- The second ANOVA was conducted with the traffic density as between factor and the situation as within factor. Significant results were revealed for both main factors, Traffic Density and Situation but not for their interaction.
- The group with traffic CAD20 showed significantly higher lateral accelerations compared to the group with no traffic (Table 5.10).
- Concerning the significant results for situation, pairwise comparisons (Holm-Bonferroni-corrected) were conducted. Results underlined the pairwise comparisons of the first ANOVA and showed that all three situations were significantly different. The crash site showed significantly higher lateral accelerations compared to the construction site (p < .001) and the interstate crossing (p < .001). In addition, the interstate crossing showed significantly higher lateral accelerations compared to the construction site (p = .03).
- No significant results were found for the factors Level of Automation, Duration and the surrogate factor Trial.

5.3.2 Subjective ratings of the take-overs

The subjective ratings of the take-overs after each situation were analyzed analogue to the objective measures to allow an assessment of potential interactions between the factors. The factor Duration was not analyzed since participants were asked to rate how they perceived the most recent take-over in the corresponding situation. They were not given

Table 5.12: Overview of the group and situation means (SDs) for the subjective ratings for the different situations.

Within factor	CAD0	CAD20	Manual20
	Subjective criticality [ ]		
Crash site	M = .95 (2.50)	M = 2.37 (1.61)	M = 2.72 (2.02)
Construction site	M = -2.25 (2.17)	M = -.84 (2.14)	M = -.39 (3.31)
Interstate crossing	M = -1.95 (2.11)	M = -.95 (3.41)	M = -.06 (2.69)
	Subjective complexity [ ]		
Crash site	M = -.75 (2.20)	M = .84 (1.46)	M = 1.28 (2.11)
Construction site	M = -2.65 (1.63)	M = -1.37 (1.92)	M = -.78 (2.86)
Interstate crossing	M = -2.00 (2.03)	M = -.21 (2.51)	M = -.39 (2.64)
	Subjective comfort [ ]		
Crash site	M = 1.40 (2.62)	M = .26 (2.92)	-
Construction site	M = 2.40 (2.16)	M = 1.16 (2.83)	-
Interstate crossing	M = 1.45 (2.95)	M = .37 (3.27)	-
	Subjective time budget [ ]		
Crash site	M = .85 (2.72)	M = -.89 (2.13)	M = .61 (2.85)
Construction site	M = 1.75 (2.27)	M = 1.84 (2.01)	M = 1.94 (3.23)
Interstate crossing	M = .60 (2.91)	M = -.79 (3.38)	M = -.72 (2.47)

a specific reference to also regard the interval of CAD prior to the take-over. Before the analysis of ANOVAs, normal distribution and the homogeneity of variance for the groups were assessed and regarded for the discussion of results. Results can be found in Table A.5. The individual group means and SDs of the subjective ratings can be found in Table 5.12. The analysis amounted to two ANOVAs per subjective rating, one for the between factor Automation Level and one for the between factor Traffic Density, both in combination with the within factor Situation. Results from the analysis can be found in Table 5.13 and the key findings are summarized in lists.

Subjective criticality

- The first ANOVA showed highly significant results for the factor Situation, but not for the Automation Level and the interaction.
- The follow-up pairwise comparisons (Holm-Bonferroni-corrected) revealed that the situation crash site was rated significantly more critical in comparison with the construction site ($p < .001$) and the interstate crossing ($p < .001$) (Figure 5.9).
- The second ANOVA for subjective criticality showed significant results for both main effects, Traffic Density and Situation. The CAD20 group rated the situations significantly more critical in comparison to the CAD0 group. For the factor situation, pairwise comparisons (Holm-Bonferroni-corrected) corroborated the results from the first ANOVA.

Table 5.13: Results from the ANOVAs conducted for the subjective ratings. If sphericity was violated, a Greenhouse-Geisser-correction was implemented and is indicated by (GG) succeeding the test statistic.

Factor combination	Main effect 1	Main effect 2	Interaction 1 x 2
	Subjective criticality		
Automation level (1) and situation (2)	$F(1,35) = 1.03$, $p = .32, \eta^2 = .01$	$F(1.98, 69.17) =$ $= 20.10$, (GG), $p < .001, \eta^2 = .25$	$F(1.98, 69.17) =$ $= .13$, (GG), $p = .88, \eta^2 < .01$
Traffic density (1) and situation (2)	$F(1,37) = 5.67$, $p = .02, \eta^2 = .07$	$F(1.84, 68.06) =$ $= 29.89$, (GG), $p < .001, \eta^2 = .29$	$F(1.84, 68.06) =$ $= .13$, (GG), $p = .87, \eta^2 < .01$
	Subjective complexity		
Automation level (1) and situation (2)	$F(1,35) = .36$, $p = .55, \eta^2 < .01$	$F(1.85, 64.84) =$ $= 8.93$, (GG), $p < .001, \eta^2 = .14$	$F(1.85, 64.84) =$ $= .32$, (GG), $p = .71, \eta^2 < .01$
Traffic density (1) and situation (2)	$F(1,37) = 12.68$, $p = .001, \eta^2 = .14$	$F(1.96, 72.68) =$ $= 13.18$, (GG), $p < .001, \eta^2 = .16$	$F(1.96, 72.68) =$ $= .20$, (GG), $p = .81, \eta^2 < .01$
	Subjective comfort		
Traffic density (1) and situation (2)	$F(1,37) = 2.47$, $p = .12, \eta^2 = .04$	$F(1.93, 71.35) =$ $= 2.72$, (GG), $p = .07, \eta^2 = .02$	$F(1.93, 71.35) =$ $= .02$, (GG), $p = .98, \eta^2 < .01$
	Subjective time budget		
Automation level (1) and situation (2)	$F(1,34) = .82$, $p = .37, \eta^2 = .01$	$F(1.92, 65.35) =$ $= 11.82$, (GG), $p < .001, \eta^2 = .15$	$F(1.92, 65.35) =$ $= 1.22$, (GG), $p = .30, \eta^2 = .02$
Traffic density (1) and situation (2)	$F(1,37) = 2.36$, $p = .13, \eta^2 = .04$	$F(1.61, 59.45) =$ $= 11.55$, (GG), $p < .001, \eta^2 = .11$	$F(1.61, 59.45) =$ $= 2.39$, (GG), $p = .11, \eta^2 = .02$

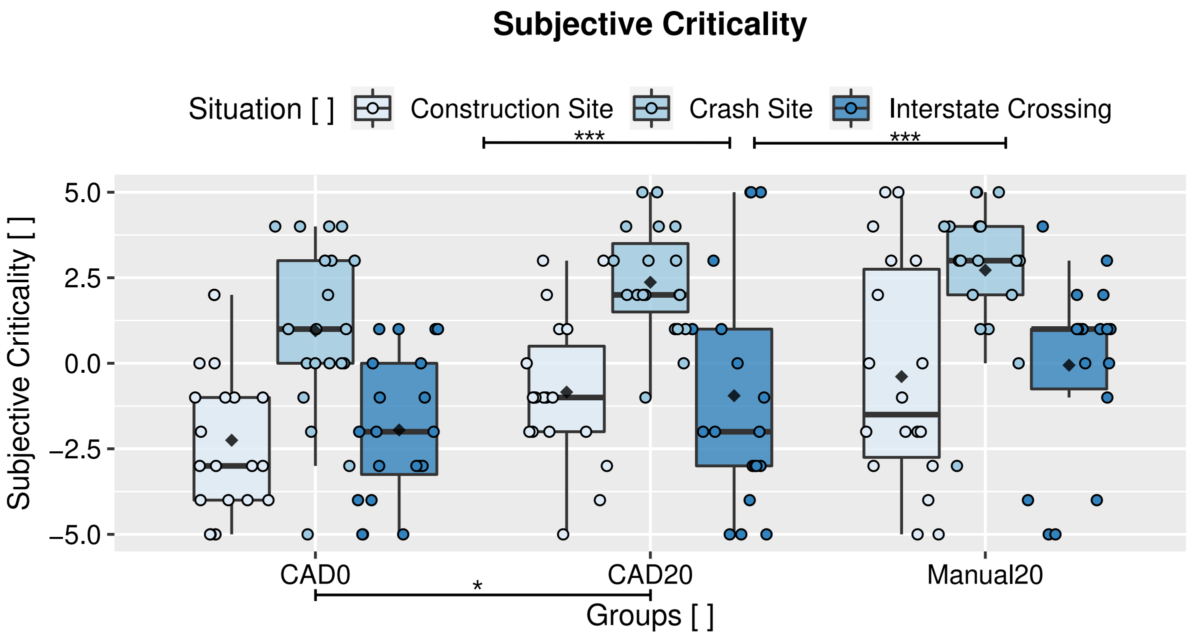


Figure 5.9: Plot of the ratings of the perceived subjective criticality of the take-overs. n_{CAD0} = 60, n_{CAD20} = 57, $n_{Manual20}$ = 54.

Subjective complexity

- Results for the subjective complexity also showed a significant difference (Holm-Bonferroni-corrected) between the crash site and the construction site ($p < .001$) and the crash site and the interstate crossing ($p = .02$) with the crash being rated more complex (Figure 5.10). The factor Automation Level and the interaction did not reveal significant results.
- The second ANOVA confirmed the results for the factor Situation with the crash site receiving the rating to be the most complex situation in comparison with the construction site ($p < .001$) and the interstate crossing ($p = .01$). The interstate crossing was rated significantly more complex compared to the construction site ($p = .03$) (Figure 5.10). All pairwise comparisons were Holm-Bonferroni-corrected.
- The factor Traffic Density also showed highly significant results with the group CAD20 rating the situations to be more complex compared to the CAD0 group (Figure 5.10). The interaction between both factors did not show significant results.

The subjective comfort was only assessed for the groups with CAD since the manual drivers did not experience a take-over whereas the subjective criticality, complexity and time budget addressed the experienced situation and not explicitly the take-over.

Subjective comfort

- Results for the subjective comfort of the take-over showed a tendency for the factor Situation and no significant results for Traffic and the interaction between them (Figure 5.11). No follow-up tests were conducted.

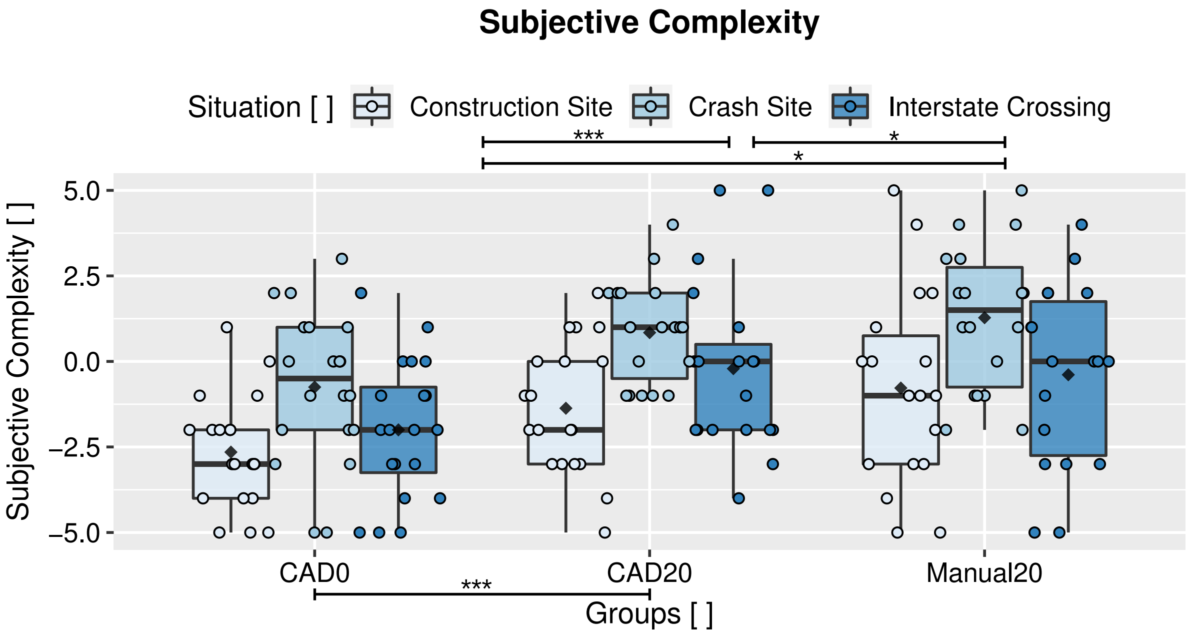


Figure 5.10: Plot of the ratings of the perceived subjective complexity of the take-overs. n_{CAD0} = 60, n_{CAD20} = 57, $n_{Manual20}$ = 54.

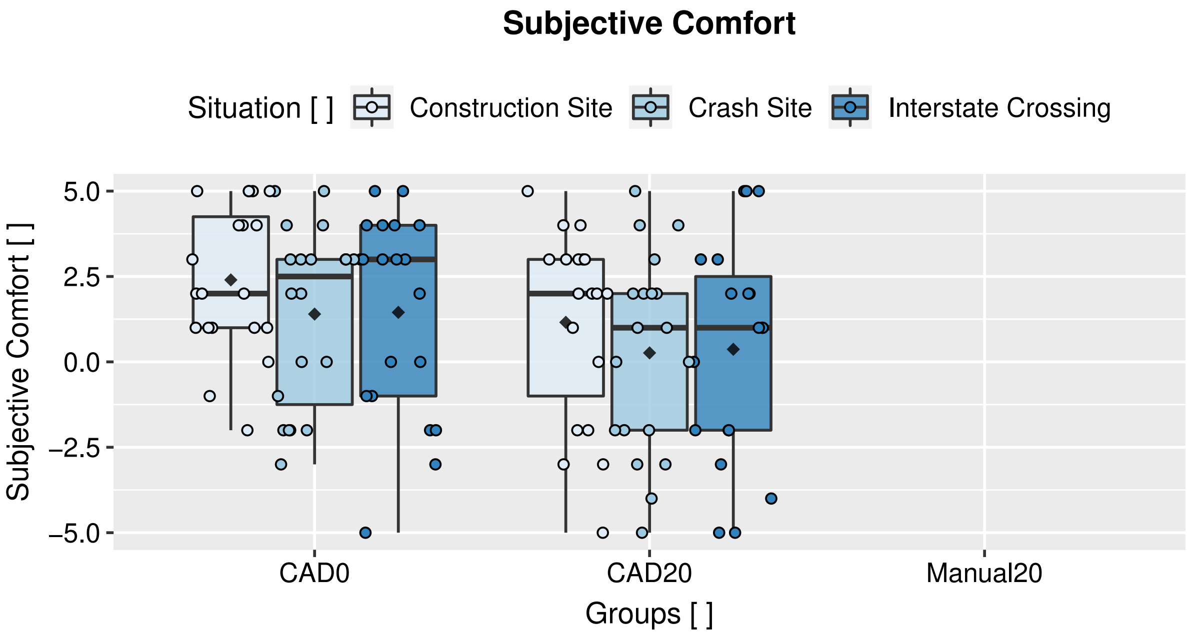


Figure 5.11: Plot of the ratings of perceived subjective comfort of the take-over. n_{CAD0} = 60, n_{CAD20} = 57.

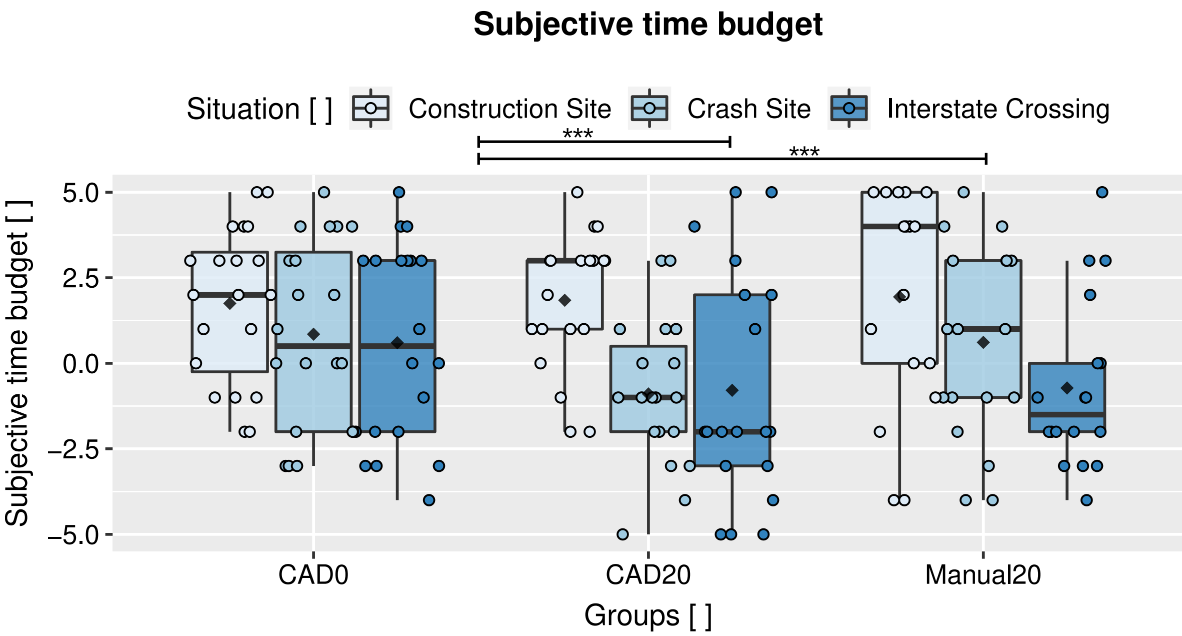


Figure 5.12: Plot of the ratings of the perceived time budget of the take-overs. $n_{CAD0} = 60$, $n_{CAD20} = 57$, $n_{Manual20} = 53$.

Subjective time budget

- Results for the subjective time budget revealed a significant result for the factor Situation but no significant effects for Automation Level and the interaction (Figure 5.12). Pairwise comparisons (Holm-Bonferroni-corrected) revealed that participants rated the construction site to exert significantly less time pressure compared to the crash site ($p = .001$) and the interstate crossing ($p < .001$).
- The second ANOVA also showed significant results for the factor Situation but not for the Traffic Density and the interaction. The pairwise comparisons (Holm-Bonferroni-corrected) underlined the significant findings from the first ANOVA.

5.3.3 State changes during CAD – Eye-Tracking

The results from the tests on normal distribution and homogeneity of variance for all eye-tracking measures can be found in Table A.6 and are regarded for the discussion of results. Individual group means and SDs are listed in Table 5.14. The respective key findings for the individual eye-tracking measures are listed according to the procedure for the subjective ratings. The data for the HGD are visualized in Figure 5.13.

HGD

- The first ANOVA for the HGD featured the factors Group and Duration and showed no significant results for both factors and their interaction.
- The second ANOVA featured the factors Group and Trial number accordingly and also did not show any significant results for both factors and their interaction (Table 5.15).

Table 5.14: Overview of the group and situation means (SDs) for the state changes assessed with eye-tracking during CAD.

Groups	5A	5B	30
		HGD [m]	
CAD0	M = .16 (.07)	M = .18 (.04)	M = .15 (.06)
CAD20	M = .18 (.08)	M = .18 (.04)	M = .19 (.05)
		PEOR [%]	
CAD0	M = 74.93 (14.77)	M = 71.63 (17.03)	M = 69.44 (22.09)
CAD20	M = 73.99 (20.73)	M = 75.81 (16.54)	M = 70.03 (20.46)
		PERCLOS [%]	
CAD0	M = 4.52 (5.25)	M = 8.05 (9.76)	M = 14.86 (27.74)
CAD20	M = 6.79 (12.96)	M = 12.68 (27.76)	M = 14.06 (24.99)
		Blink duration [s]	
CAD0	M = .35 (.07)	M = .35 (.09)	M = .38 (.08)
CAD20	M = .38 (.08)	M = .36 (.05)	M = .37 (.07)
		Blink frequency [Hz]	
CAD0	M = .30 (.23)	M = .31 (.25)	M = .25 (.19)
CAD20	M = .24 (.18)	M = .21 (.13)	M = .21 (.15)

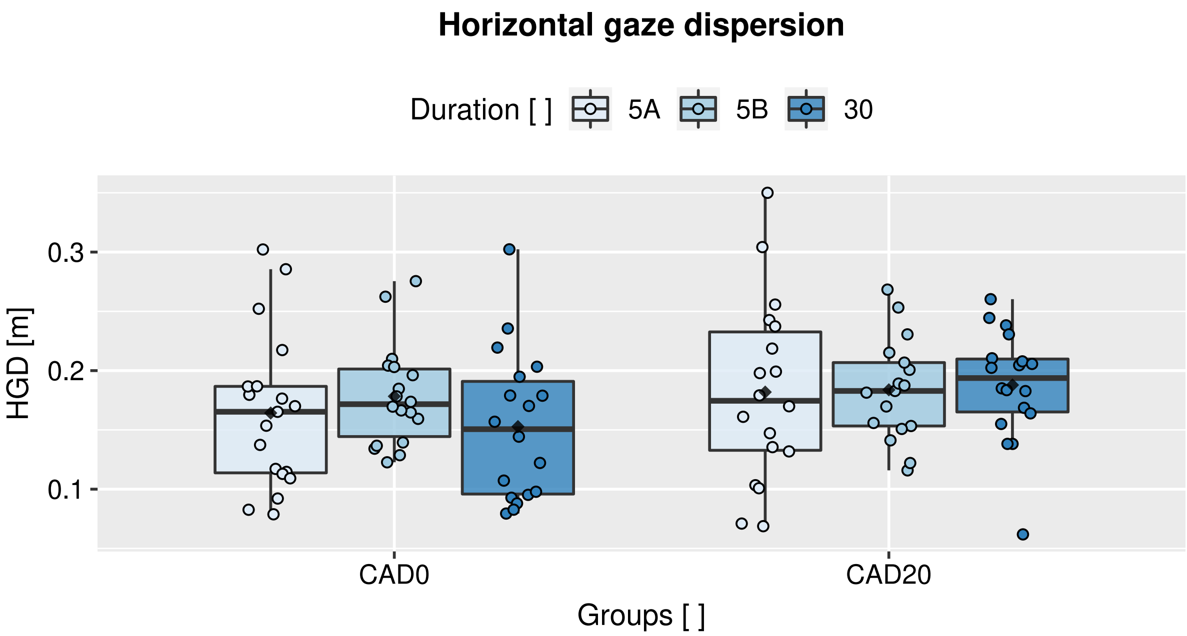


Figure 5.13: Plot of the SD of the horizontal gaze position representing the HGD for the last minute before the respective take-overs, n_{CAD0} = 55, n_{CAD20} = 53.

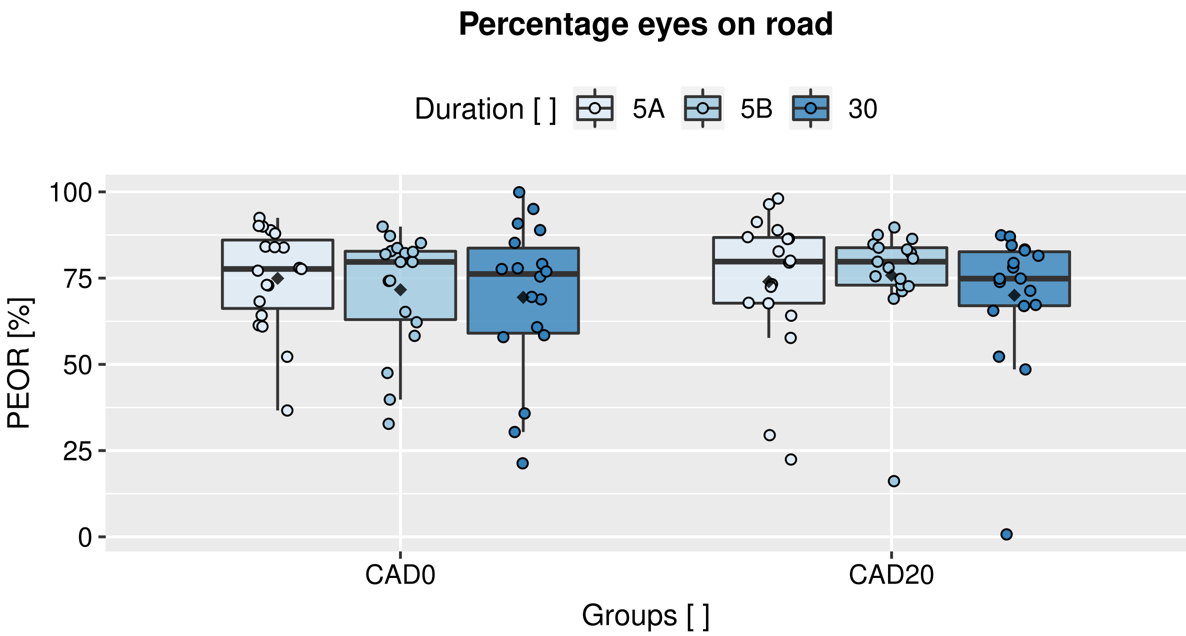


Figure 5.14: Plot of the percentage eyes on road (PEOR) for the last minute before the respective take-overs, n_{CAD0} = 55, n_{CAD20} = 53.

Since the HGD mainly represents a measure of activity, the percentage eyes on road (PEOR) during the last minute before a respective take-over were also analyzed to allow a more comprehensive understanding of drivers visual behavior. The data for the HGD are plotted in Figure 5.14.

PEOR

- Analysis for the PEOR consisted of two ANOVAs, one with the factors Group and Duration and one with the factors Group and Trial number. The first ANOVA showed no significant results for both main effects and their interaction, nor did the second ANOVA (Table 5.15).

The data for PERCLOS are visualized in Figure 5.15.

PERCLOS

- A mixed-design ANOVA was calculated for analysis of the PERCLOS. Results showed a tendency for significant results for the factor Duration but not for the factor Group and the interaction.
- Due to the p-value of .06, pairwise comparisons (Holm-Bonferroni-corrected) were conducted and showed a significant difference for the PERCLOS values of the last minute of the automated drive between the first 5-minute drive (5A) and the 30-minute automated drive (p = .04) (Figure 5.15). The PERCLOS-values were significantly higher following the pairwise comparisons at the end of the 30-minute drive compared to the first 5 minutes (5A), but not compared to the second 5-minute drive (5B).

Table 5.15: Results of the ANOVAs conducted for the state changes assessed with eye-tracking. The values represent the corresponding driver state in the last minute before a respective take-over. If sphericity was violated, a Greenhouse-Geisser-correction was implemented and is indicated by (GG) succeeding the test statistic.

Factor combination	Main effect 1	Main effect 2	Interaction 1 x 2
		HGD	
Group (1) and duration (2)	$F(1,33) = 1.07,$ $p = .31, \eta^2 = .02$	$F(1.76, 58.24) =$ $= 1.12,$ (GG), $p = .33, \eta^2 = .01$	$F(1.76, 58.24) =$ $= 1.01,$ (GG), $p = .36, \eta^2 = .01$
Group (1) and trial (2)	$F(1,33) = 1.07,$ $p = .31, \eta^2 = .02$	$F(1.96, 64.66) =$ $= .30,$ (GG), $p = .74, \eta^2 < .01$	$F(1.96, 64.66) =$ $= 1.13,$ (GG), $p = .33, \eta^2 = .01$
		PEOR	
Group (1) and duration (2)	$F(1,33) = .23,$ $p = .63, \eta^2 < .01$	$F(1.70, 56.19) =$ $= 1.64,$ (GG), $p = .21, \eta^2 = .02$	$F(1.70, 56.19) =$ $= .12,$ (GG), $p = .86, \eta^2 < .01$
Group (1) and trial (2)	$F(1,33) = .23,$ $p = .63, \eta^2 < .01$	$F(1.63, 53.63) =$ $= .29,$ (GG), $p = .70, \eta^2 < .01$	$F(1.63, 53.63) =$ $= 1.45,$ (GG), $p = .24, \eta^2 = .01$
		PERCLOS	
Group (1) and duration (2)	$F(1,32) = .15,$ $p = .70, \eta^2 < .01$	$F(1.59, 50.94) =$ $= 3.17,$ (GG), $p = .06, \eta^2 = .03$	$F(1.59, 50.94) =$ $= .45,$ (GG), $p = .59, \eta^2 < .01$
Group (1) and trial (2)	$F(1,32) = .15,$ $p = .70, \eta^2 < .01$	$F(1.28, 41.08) =$ $= 4.17,$ (GG), $p = .04, \eta^2 = .04$	$F(1.28, 41.08) =$ $= .04,$ (GG), $p = .89, \eta^2 < .01$
		Blink duration	
Group (1) and duration (2)	$F(1,32) = .31,$ $p = .58, \eta^2 < .01$	$F(1.87, 59.81) =$ $= .68,$ (GG), $p = .50, \eta^2 = .01$	$F(1.87, 59.81) =$ $= 1.18,$ (GG), $p = .31, \eta^2 = .02$
Group (1) and trial (2)	$F(1,32) = .31,$ $p = .58, \eta^2 < .01$	$F(1.80, 57.54) =$ $= 1.35,$ (GG), $p = .27, \eta^2 = .02$	$F(1.80, 57.54) =$ $= .06,$ (GG), $p = .93, \eta^2 < .01$
		Blink frequency	
Group (1) and duration (2)	$F(1,32) = .70,$ $p = .41, \eta^2 = .02$	$F(1.88, 60.12) =$ $= 1.65,$ (GG), $p = .20, \eta^2 < .01$	$F(1.88, 60.12) =$ $= .73,$ (GG), $p = .48, \eta^2 < .01$
Group (1) and trial (2)	$F(1,32) = .70,$ $p = .41, \eta^2 = .02$	$F(1.89, 60.46) =$ $= .36,$ (GG), $p = .69, \eta^2 < .01$	$F(1.89, 60.46) =$ $= .42,$ (GG), $p = .65, \eta^2 < .01$

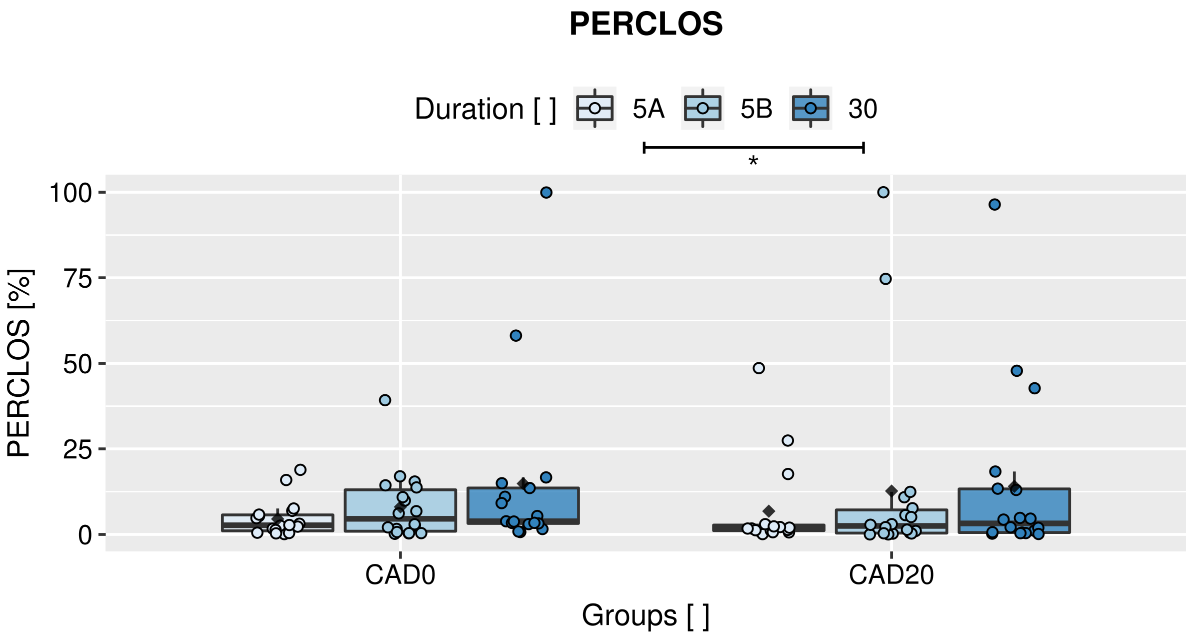


Figure 5.15: Plot of the PERCLOS values for the last minute before the respective take-overs, n_{CAD0} = 53, n_{CAD20} = 53.

- The second ANOVA analyzed the PERCLOS values of the last minute before a take-over with regard to the Trial of the take-overs. Results showed significant results for Trial Number, but not for Group and their interaction. Pairwise comparisons (Holm-Bonferroni-corrected) revealed a significant difference between PERCLOS values right before the first take-over and the third/last take-over (p = .02) (Table 5.15)

The blink duration was analyzed concerning the absolute values for the last minute prior to the respective take-overs for the individual groups and the different durations of CAD. The data for blink duration is plotted in Figure 5.16.

Blink duration

- The ANOVA for the factors Group and Duration did not reveal any significant results for both main factors and their interaction (Table 5.15).
- In addition to the analysis of previous eye-tracking measures, a second ANOVA was conducted with the factors Group and trial number to allow a detailed look on the overall effect of time. Results did not show any significant results for both factors and their interaction (Table 5.15).

Corresponding to an increase in blink duration, the blink frequency decreases respectively for drowsy drivers and was also analyzed. The data for blink frequency were plotted in Figure 5.17.

Blink frequency

- Both ANOVAs showed no significant results for the factors Group and Duration, the factors Group and Trial Number and their corresponding interactions (Table 5.15).

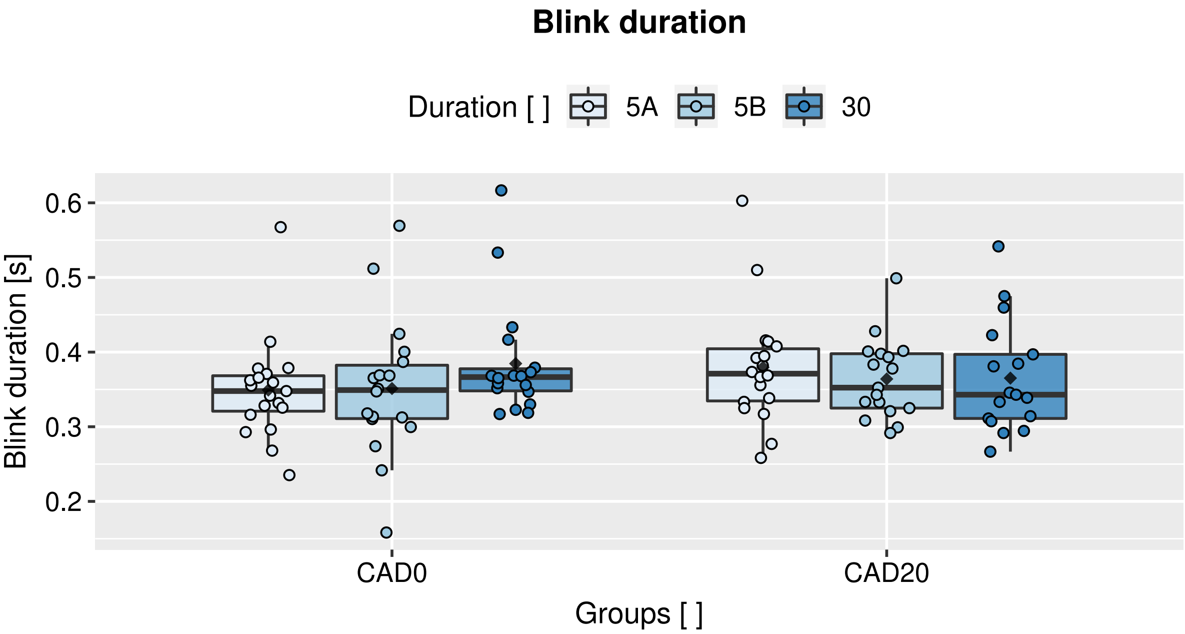


Figure 5.16: Plot of the blink duration for the last minute before the respective take-overs, n_{CAD0} = 55, n_{CAD20} = 52.

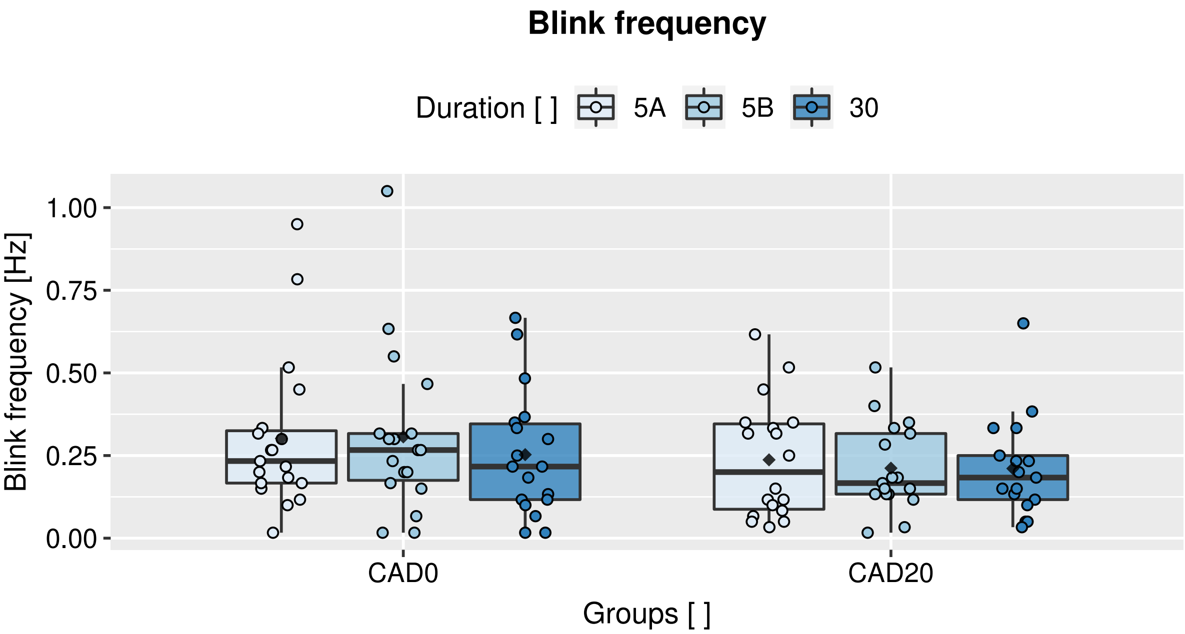


Figure 5.17: Plot of the blink frequency for the last minute before the respective take-overs, n_{CAD0} = 55, n_{CAD20} = 52.

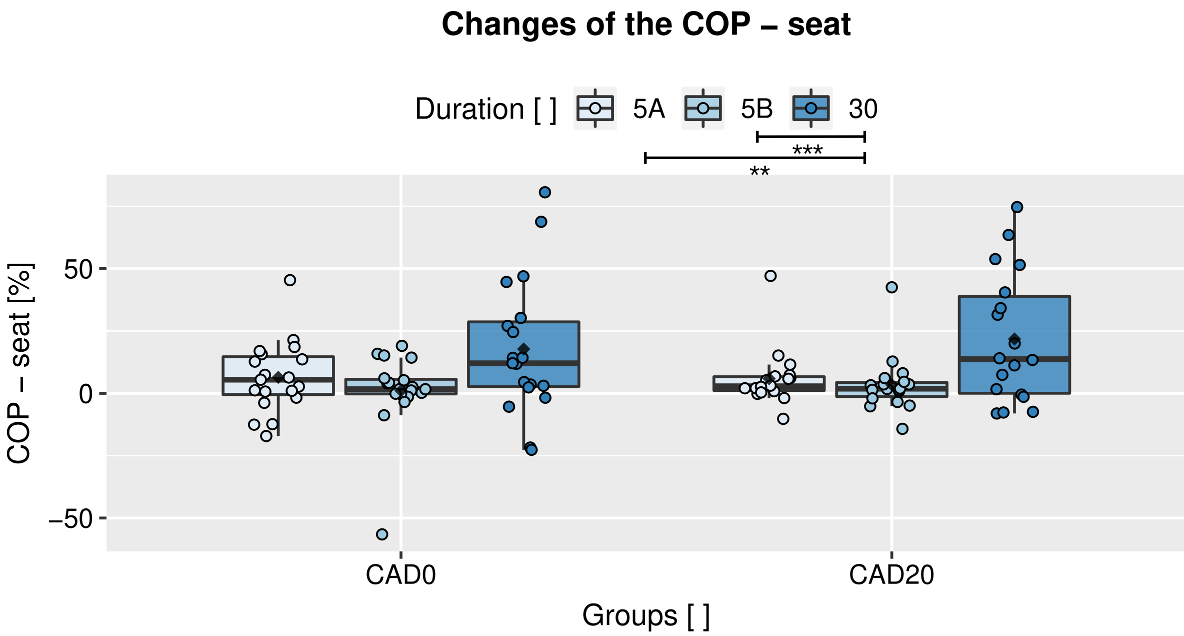


Figure 5.18: Plot of the changes of the COP in the seat between the second minute and last minute of the respective CAD durations, $n_{CAD0} = 57$, $n_{CAD20} = 54$.

5.3.4 State changes during CAD – Seat pressure mats

The analysis of the center of pressure (COP) was conducted for the seat and back rest and follows the general approach of calculating the COP and respective changes depicted in Chapter 4.3. Results on the data distribution and homogeneity of variance can be found in Table A.7 and will be regarded in the discussion of the results. The individual group means and SDs can be found in Table 5.16 and the key findings are summarized in lists.

COP – seat

- With regard to the previous analysis of measures, an ANOVA was conducted with the factors Group and Duration. Results showed no significant results for the factor Group and the interaction, but highly significant results for the factor Duration. Pairwise comparisons (Holm-Bonferroni-corrected) were conducted and showed significant differences between the 30-minute automated drive and the first 5-minute drive ($p < .01$) and the second 5-minute drive ($p < .001$), but no significant differences between the two 5-minute drives (Figure 5.18). The changes of the COP in the seat became significantly larger in the 30-minute automated drive compared to the 5-minute drives.

- A second ANOVA was conducted to assess the overall effect of time, utilizing the factors Group and Trial number. Results showed no significant results for the factors Group, Trial Number and their interaction.

Corresponding to the changes of the COP in the seat, the changes of the COP in the backrest were also assessed. The plotted data can be seen in Figure 5.19.

COP – backrest

Table 5.16: Overview of the group and situation means (SDs) for the state changes assessed with the seat pressure mats during CAD.

Groups	5A	5B	30
		COP – seat [%]	
CAD0	M = 6.40 (14.38)	M = 1.08 (15.68)	M = 17.75 (27.34)
CAD20	M = 5.88 (11.62)	M = 3.44 (11.37)	M = 21.79 (26.06)
		COP – backrest [%]	
CAD0	M = .57 (13.68)	M = 6.24 (15.55)	M = 15.95 (27.41)
CAD20	M = -.46 (10.13)	M = 1.32 (13.07)	M = 17.02 (33.00)
		Contact area – seat [%]	
CAD0	M = 17.70 (88.75)	M = -8.35 (59.98)	M = -5.21 (63.90)
CAD20	M = 8.89 (37.12)	M = 19.60 (88.59)	M = 26.84 (93.74)
		Contact area – backrest [%]	
CAD0	M = 2.13 (76.04)	M = -6.40 (86.71)	M = 54.08 (139.31)
CAD20	M = 31.69 (66.30)	M = 22.08 (123.16)	M = 52.84 (101.98)

- The first ANOVA showed significant results for the factor Duration but not for the factor Group and their interaction. Succeeding pairwise comparisons (Holm-Bonferroni-corrected) between the different durations revealed significant differences between the 30-minute automated drive with the 5A-drive ($p < .001$) and the 5B-drive ($p < .01$) but not between the two 5-minute drives. The changes of the COP in the backrest were significantly larger for 30 minutes of CAD compared to 5 minutes of CAD.
- The second ANOVA with factors Group and Trial number showed no significant results for both main factors and their interaction.

The activity of participants in the seat and backrest was additionally assessed by looking at the changes of the contact area following the depiction in Chapter 4.3. The activity of drivers during CAD was analyzed utilizing the change of the variance of the contact between the second and last minute of the different CAD durations. Data were checked concerning the distribution and homogeneity of variance and results can be found in Table A.7. The group means and SDs are in Table 5.16. In a first step, the changes of the contact area of the seat were analyzed. The corresponding data are plotted in Figure 5.20 and the key findings are listed.

Contact Area – seat

- The first ANOVA featured the factors Group and Duration and showed no significant results for both factors and their interaction nor did the second ANOVA with the factors Group and Trial number (Table 5.17).

The changes of the variance of the contact area in the backrest were also analyzed. Figure 5.21 shows the plotted data.

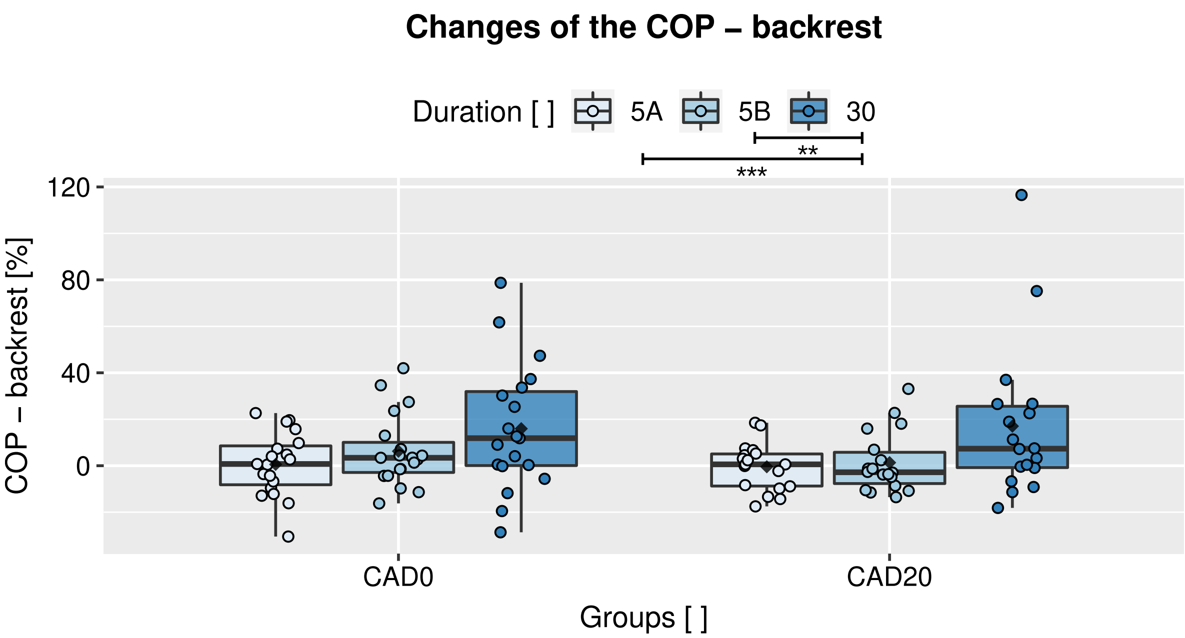


Figure 5.19: Plot of the changes of the COP in the backrest between the second minute and last minute of the respective CAD durations, n_{CAD0} = 57, n_{CAD20} = 54.

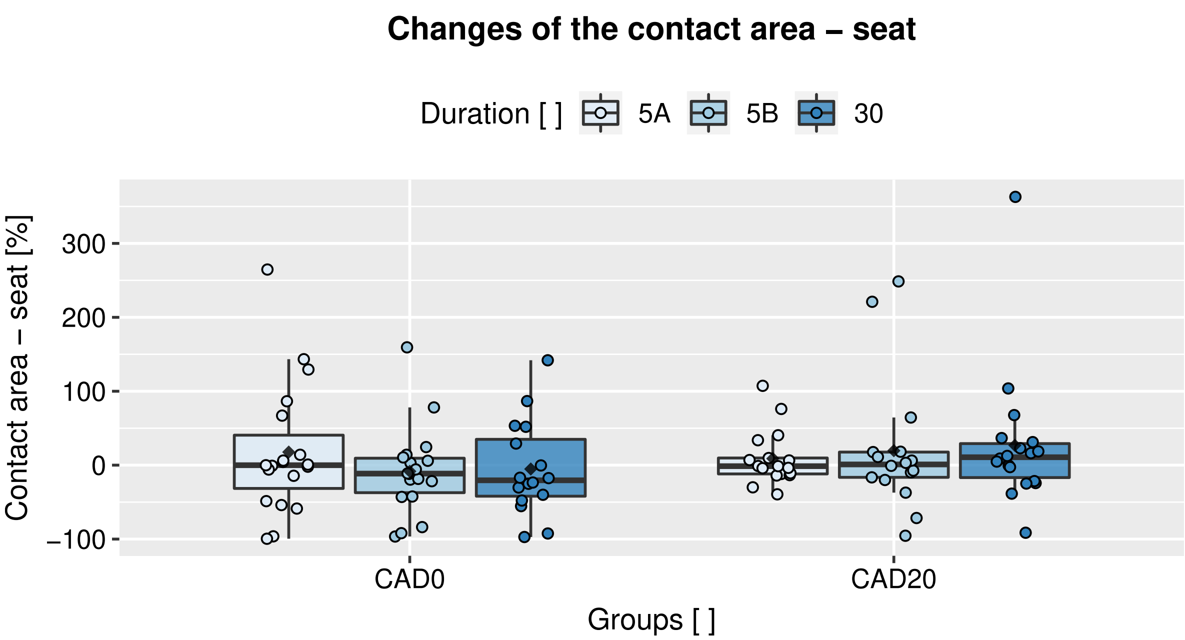


Figure 5.20: Plot of the changes of the variance of the contact area in the seat between the second minute and last minute of the respective CAD durations, n_{CAD0} = 53, n_{CAD20} = 52.

Table 5.17: Results from the ANOVAs conducted for the state changes recorded in the seat pressure mats. The values represent the corresponding driver state between the second and last minute of the respective automated drive durations. If sphericity was violated, a Greenhouse-Geisser-correction was implemented and is indicated by (GG) succeeding the test statistic.

Factor combination	Main effect 1	Main effect 2	Interaction 1 x 2
		COP – seat	
Group (1) and duration (2)	$F(1, 35) = .28$, $p = .60, \eta^2 < .01$	$F(1.67, 58.52) =$ $= 8.95$, (GG), $p < .001, \eta^2 = .14$	$F(1.67, 58.52) =$ $= .14$, (GG), $p = .83, \eta^2 < .01$
Group (1) and trial (2)	$F(1, 35) = .28$, $p = .60, \eta^2 < .01$	$F(1.92, 67.29) =$ $= 2.72$, (GG), $p = .08, \eta^2 = .05$	$F(1.92, 67.29) =$ $= .16$, (GG), $p = .85, \eta^2 < .01$
		COP – backrest	
Group (1) and duration (2)	$F(1, 35) = .12$, $p = .73, \eta^2 < .01$	$F(1.54, 53.82) =$ $= 8.25$, (GG), $p < .01, \eta^2 = .11$	$F(1.54, 53.82) =$ $= .26$, (GG), $p = .72, \eta^2 < .01$
Group (1) and trial (2)	$F(1, 35) = .12$, $p = .73, \eta^2 < .01$	$F(1.90, 66.36) =$ $= .47$, (GG), $p = .62, \eta^2 < .01$	$F(1.90, 66.36) =$ $= 2.52$, (GG), $p = .09, \eta^2 = .04$
		Contact area – seat	
Group (1) and duration (2)	$F(1, 29) = .51$, $p = .48, \eta^2 < .01$	$F(1.97, 57.26) =$ $= .77$, (GG), $p = .46, \eta^2 = .02$	$F(1.97, 57.26) =$ $= 1.20$, (GG), $p = .31, \eta^2 = .03$
Group (1) and trial (2)	$F(1, 29) = .51$, $p = .48, \eta^2 < .01$	$F(1.77, 51.40) =$ $= .67$, (GG), $p = .50, \eta^2 = .02$	$F(1.77, 51.40) =$ $= .86$, (GG), $p = .42, \eta^2 = .02$
		Contact area – backrest	
Group (1) and duration (2)	$F(1, 31) = .58$, $p = .45, \eta^2 < .01$	$F(1.98, 61.43) =$ $= 2.13$, (GG), $p = .13, \eta^2 = .04$	$F(1.98, 61.43) =$ $= .55$, (GG), $p = .58, \eta^2 = .01$
Group (1) and trial (2)	$F(1, 31) = .58$, $p = .45, \eta^2 < .01$	$F(1.89, 58.67) =$ $= 2.00$, (GG), $p = .15, \eta^2 = .04$	$F(1.89, 58.67) =$ $= .11$, (GG), $p = .89, \eta^2 < .01$

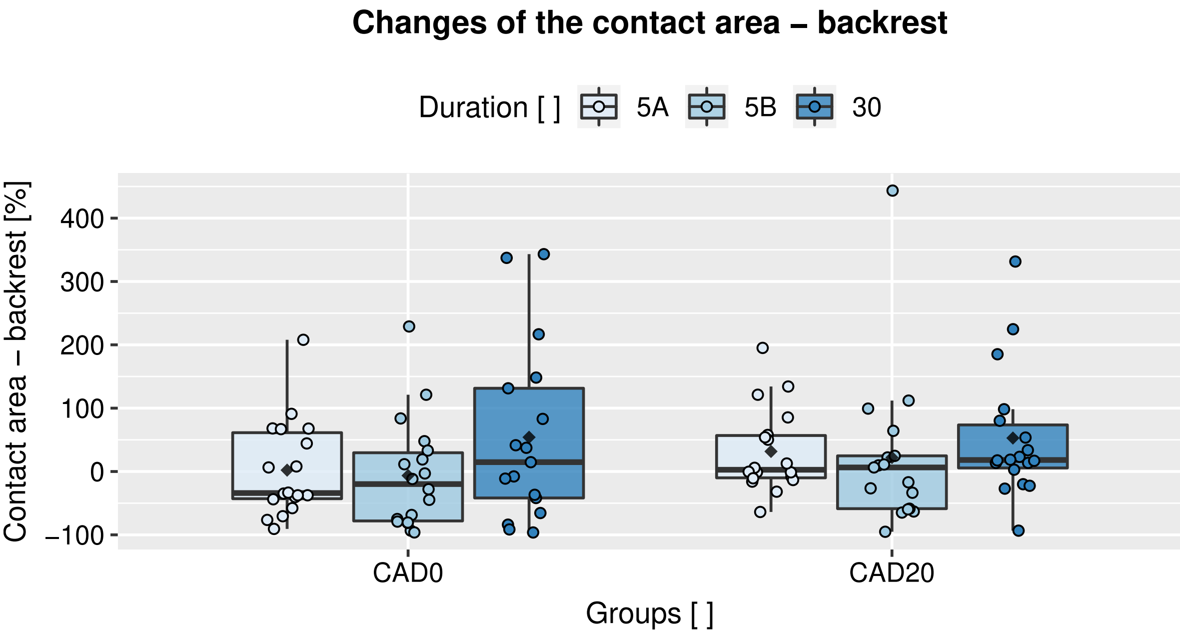


Figure 5.21: Plot of the changes of the variance of the contact area in the backrest between the second minute and last minute of the respective CAD durations, n_{CAD0} = 53, n_{CAD20} = 53.

Contact Area – backrest

- The first ANOVA assessed the factors Group and Duration and showed no significant results for both factors and their interaction nor did the second ANOVA (Table 5.17).

5.4 Discussion

Data from this experiment provide a broad empirical foundation for the assessment of prolonged CAD on the driver state, the take-over performance in different situations and the subjective rating of take-overs. In addition, manual drivers experiencing these situations allow a comparison of take-over behavior to the current status quo concerning various measures.

Overall, the data from this experiment featured predominantly non-normal distributions, potentially impeding the clear analysis of underlying effects. Homogeneity of Variance was not violated in a way leading to an adjustment of the p-value. Significant interaction results from various ANOVAs underline the feasibility of assessing the data with ANOVAs regardless of the non-normal distributions. The quality aspects of take-over performance, e.g. accelerations and the TTC showed highly significant interaction results, shedding light on the need for a differentiated analysis of underlying effects.

The discussion is aligned with the research questions in Figure 5.1 and provides answers based on the empirical findings that are put in critical perspective to the literature findings discussed in chapter 2. The first research question RQ1 focused on the effect of prolonged, monotonous CAD on the driver state and if the progression is dependent on different traffic densities. Results of driver state measures from both eye-tracking and the seat pressure mats showed no significant results for the HGD, PEOR and blink duration and frequency. However PERCLOS, as most prominent measure concerning drowsiness (see Chapter

2.2.3), revealed both small effects for duration and trial. Participants were significantly more drowsy after 30 minutes of CAD compared to 5 minutes of CAD, but only if the 5-minute-interval came before the 30-minute-interval. With regard to the significant results concerning trial, prolonged, monotonous CAD (30 to 40 minutes) led to a minor onset of drowsiness. This effect did not affect the visual tracking behavior (HGD, PEOR) but was also visible in the seat pressure mats. Compared to the respective first minutes of CAD, participants showed more unrest before the take-over in the 30-minute-interval. While this could link to the onset of drowsiness, it can also be argued that prolonged CAD leads to an increase in seating discomfort which is met by shifting positions more often. Both the seat and the backrest showed significant differences between the 30-minute interval and both 5A- and 5B-intervals, relining the argument, that the seat pressure mats did not capture the onset of drowsiness, but a different effect, potentially linked to seating comfort. Since there were no significant differences between the two five-minute intervals in all measures of driver state, the action of taking over appears to temporarily increase the level of arousal.

In line with literature findings, showing increased drowsiness after prolonged CAD (J. Gonçalves et al., 2016; Vogelpohl, Kühn, Hummel, & Vollrath, 2018; Bourrelly et al., 2019), PERCLOS can be recommended as favorable measure for drowsiness. Seat pressure mats are able to capture a potential additional effect from prolonged CAD, but should not be argued as resilient measure for e.g. drowsiness. Future work could more heavily rely on a design of experiment featuring a fixed level of drowsiness rather than a fixed time prior to the take-over to magnify effects from increased drowsiness. While this provides chances concerning the effect of drowsiness, potential other effects, apparent in this case in the seat pressure mats, would not be regarded. Thus, future work should rather focus on capturing all relevant changes of the driver state compared to a content-related narrow but bloated effort centering on drowsiness.

In contrast to the results for prolonged CAD, none of the eye-tracking and seat pressure mats measures showed significant results for the effect of traffic density. Based on this work, traffic density does not affect the progression of the driver state in CAD. Since NDRTs are allowed, direct visual attention on the traffic situation during active CAD seems unlikely in future use cases.

The research question RQ3 addressed the effect of potential driver state changes on the take-over performance. No significant effects were apparent for duration and trial for TOT, TTC, and longitudinal and lateral accelerations. While minor state changes were visible, RQ3 and the question of how these affect take-over performance must be dismissed on the grounds of not seeing any effect on take-over performance. The small magnitude of detectable state changes could be the reason for the findings while they underline publications not showing effects from increased drowsiness (J. Schmidt et al., 2016, 2017). Taking into account publications showing minor changes concerning quality aspects of take-over performance (Kreuzmair et al., 2017; Feldhütter, Kroll, & Bengler, 2018), a reasonable explanation for the dismissal of RQ3 is tying into RQ4 and the effect of different situations. Throughout results of TOT, TTC and accelerations, highly significant differences between traffic density and the take-over situations were found. The magnitude of effects from different situations can be interpreted as to overshadow state changes. The start of a take-over situation with issuing a RtI appears to level the changes originating from prolonged CAD. Following the framework in Figure 2.3 (Marberger et al., 2017), both motoric state changes (COP) as well as arousal level changes (PERCLOS) were compensated by participants in the driver state transition process before intervention.

Contrary, the requirements of the take-over scenario must be regarded in great detail. The variety of interaction and main effects is interpreted based on the process depicted in Chapter 4.3. With regard to the overall criticality rating of the situations following the methodology presented in Chapter 4.1, the construction site represents the situation with the lowest overall criticality, especially without additional traffic in the CAD0-group.

Up to a certain limit set by the time budget and moderating parameters such as traffic density or the complexity of the necessary driver reaction, a higher situational criticality motivated faster reactions in the interstate crossing. This effect is mitigated by adding further urgency or criticality to the point were TOTs cannot be reduced but the quality parameters of the take-over decrease drastically in highly critical and complex situations. Considering the effect sizes, different situations appear to affect the TOT in a small to moderate way, in line with the literature (Gold, 2016; Roche & Brandenburg, 2018).

The interaction results between traffic density and situation showed large effect sizes, corroborating the delicate design of take-over situations: in the less critical construction site, no traffic led to smaller TTC-values, since participants took more time to take-over, whereas additional traffic promoted more critical reactions in the interstate crossing and the crash site. Results from the longitudinal acceleration underline this interpretation, by also showing significant interaction results. Participants greatly differed in their braking behavior depending on the traffic density in otherwise identical situations. The results from the lateral acceleration do not show significant interaction results, but show highly significant differences between the situations and traffic density.

An overall criticality assessment of take-over behavior based on TTC and accelerations is highly dependent on the traffic situation at hand. Large effect sizes in combination with a highly significant effect of traffic density call for a critical view on issuing a RtI simply based on detecting a system limit. While CAD implies fallback-ready users in any situation (SAE J3016, 2018), the empirical evidence suggests to gather basic information a priori on the specific take-over situation to allow a feasible prediction of take-over performance.

Results for the subjective criticality, complexity and time budget underline the findings from the objective take-over performance, showing highly significant situational differences with large effect sizes in addition to an effect of traffic density. Surprisingly, while participants rated the situations and the traffic density differently, the subjective comfort of the take-over showed no significant difference but a high variance. Participants either had an individually different understanding of take-over comfort or rated the comfort of a take-over independent of the criticality, complexity and the subjective time budget. A full assessment of the applicability of CAD must include objective measures as well as subjective ratings. The data from this experiment support the conclusion that drivers in CAD are both able to and willing to act as fallback ready users looking at objective and subjective measures of take-over performance.

In comparison to manual drivers experiencing identical situations, drivers taking over only revealed significantly lower TTCs, but no significant differences for longitudinal and lateral accelerations. While lower TTCs represent more critical maneuvers, no significant differences for the accelerations put this reading in perspective. CAD requires driver to react to the RtI and take-over, having less time to execute similar maneuvers compared to manual drivers being in the loop, leading to smaller TTCs. Thus, take-overs should not be labeled to be more critical than manual driver reactions due to a lack of indication concerning acceleration results. More likely, TOT by definition reduces the time available for the interaction in a take-over, leading to the TTC results. Including literature findings, the take-over performance in CAD is similar to maneuvers from manual drivers (Skottke et al., 2014). Effects for the TTC are attributed to less time in the situation compared to manual drivers, but remain small compared to situational influences found in this work and literature findings (Skottke et al., 2014).

5.5 Summary and conclusion

Concluding, different take-over situations, interlinked with an effect of traffic density, are unequivocally affecting both TOT and take-over quality measures. While these effects are corroborated by the subjective rating from participants, they show large effect sizes and are in line with previous research. Effects from prolonged periods of CAD can be seen analyzing specific measures such as PERCLOS or the changes in the COP in this work. These changes in the driver state show small effect sizes and are not manifested by assessing additional measures such as blink duration or blink frequency. More importantly, this experiment did not show any effects of prolonged CAD on take-over performance whereas large situational differences are apparent. The effect of CAD assessed in take-over situations and compared to manual driving did not show differences in measures of take-over performance with the exception of the TTC.

Future research should address which state changes possess the possibility of affecting take-over performance while allowing a reasonable assessment of take-over behavior in overall less critical situations. In addition, individual differences were not regarded in this analysis but incorporate a deeper insight in relevant effects on take-over performance (Radlmayr, Feldhütter, et al., 2018).

6 The effects of non-driving related tasks on driver state and take-over performance

This experiment[1] focused on the effects of different NDRTs on the development of the driver state and the ensuing effects on take-over performance in different situations. The experiment was pre-published to this thesis at the 20th Congress of the International Ergonomics Association (IEA) in Florence (Radlmayr, Fischer, & Bengler, 2019). This chapter provides a comprehensive overview of the experiment and the most important findings and is taken in parts from the publication. For a more detailed depiction refer to Radlmayr, Fischer, and Bengler (2019).

Various effects from different NDRTs on take-over performance have already been studied (see Chapter 2.2.2). Radlmayr et al. (2014) found similar effects on take-over performance between the visual surrogate reference task (SuRT) and the cognitive n-back task with an overall higher number of crashes for the SuRT in the most critical situation. Based on the literature review, the research questions depicted in Figure 6.1 were derived and aim to analyze the development of the driver state depending on different NDRTs in more detail. Analogue to Experiment 1, potential changes are assessed in different take-over situations and allow an assessment of the effect of standardized NDRTs in addition with taking into account a variable instruction of participants. In this experiment, a total of three different NDRTs, the SuRT, the n-back task and a motoric task, a shape sorter ball hidden from view in a fabric bag were evaluated. The tasks were chosen to represent standardized tasks applying to specific modalities and their potential interaction with take-overs following the multiple resource theory (Wickens & Liu, 1988). The SuRT represents a mainly visual and motoric tasks (ISO/TS 14198, 2012). The n-back task represents the cognitive task (Kirchner, 1958) which was also validated as working memory measure (Jaeggi, Buschkuehl, Perrig, & Meier, 2010). The shape sorter ball representing a mainly motoric task is based on work from Gold et al. (2015). Figure 6.2 shows the SuRT as implemented in the car and the shaper sorter ball that was hidden in a bag during the experiment. The instruction to engage in a specific task was either instructed or participants could chose freely to engage in one of the three tasks or not engage in one of

Driver State

Experiment 2: Effect of NDRTs

Pre-published in Radlmayr, Fischer, and Bengler (2019).

RQ2 How do different NDRTs affect the driver state?
   a) Is there a difference in the engagement in NDRTs depending on the instruction?

RQ3 How do potential driver state changes affect the take-over performance?

RQ4 How does the effect from driver state changes on take-over performance compare to the effect of different situations?

Chapter 6

Figure 6.1: Summary of the research questions for Experiment 2.

[1]The experiment was designed and conducted with the assistance of Fabian Marco Fischer as part of his master's thesis (Fischer, 2016).

Figure 6.2: The left figure shows the SuRT in the center console of the vehicle. The right figure shows the shape-sorter ball that was put into a fabric bag with openings during the experiment to only allow manual (and not visual) interaction during the task. Figures with friendly permission by Fabian Marco Fischer.

the tasks at all. Take-over performance was evaluated in two different take-over situations known from Chapter 4 differing in their overall criticality. The crash site represented a critical situation and the construction site represented medium overall criticality. The design of experiment consisted of a mixed design, with the factor Type of NDRT (no, SuRT, n-back task, motoric task) being a between factor, leading to three groups with the within-factors Instruction (instructed, free) and Situation (crash site, construction site). To counterbalance the procedure, participants experienced both situations twice, leading to a total of four take-overs per participant. The factors were all permuted to avoid learning effects.

Fifty-tree participants were part of the experiment in the static simulator at the Chair of Ergonomics at the Technical University of Munich. For more details on the general experimental setup refer to Chapter 4. For data acquisition of driver state changes, an eye-tracker and seat pressure mats were used. Most important dependent variables to assess driver state were PEOR, standard deviation of the horizontal gaze position (HGD), blink frequency and changes in the COP and contact area in the seat and backrest. To analyze take-over performance, TOT, TTC, standard deviation of lateral position (SDLP), longitudinal and lateral accelerations and subjective ratings were evaluated.

> "Results show that the use of eye-tracking and seat-pressure mats allows the detection of changes in driver availability to some extent. The HGD cannot be used to differentiate between different modalities, but allows the detection of engagement into a NDRT in general. Blink frequency also shows significant changes between the NDRTs but also the situations. This either shows the influence of the track or more likely the large individual differences between participants. The significant results for the blink frequency should be viewed critically. Participants react significantly faster in the crash site situation, which can be attributed to the higher overall criticality of the "crash site" compared to the “construction site” adding to a perceived urgency (Gold, 2016). The higher criticality is punctuated by results of the TTC, the accelerations and the subjective ratings of the two situations. The lateral accelerations are within expectation, since the construction site does not feature a lane change maneuver. Results from the free behavior should be viewed critically, since standardized NDRTs were offered but no realistic tasks or activities.

> Nonetheless it can be concluded that participants take up on the offer of engaging into NDRTs in conditionally automated driving.
>
> Results from this experiment are in line with previous findings but offer an additional, more detailed assessment of changes of driver [state] during automated driving. Concluding, different modalities of NDRTs do not seem to effect take-over performance in a critical way. Contrary, the influence of different take-over situations is revealed and is consistent with findings from the overall scope of research. Eye-tracking and seat pressure mats offer a promising way of assessing changes in driver [state] even though, in this experiment, they did not result in changes of the take-over performance accordingly" (Radlmayr, Fischer, & Bengler, 2019).

Results from Experiment 1 (Chapter 5) and the findings from this experiment both showed a highly significant effect of different take-over situations on take-over performance. While eye-tracking and seat pressure mats allowed the detection of driver state changes during CAD in both experiments, no significant effect on take-over performance was detected. Individual predispositions and differences concerning the development of the driver state during CAD were not regarded in both experiments but hypothesized to effect results as well. In addition to publications strongly suggesting to incorporate individual differences and predispositions in a comprehensive understanding of relevant effects on take-over performance (Gold, 2016), data from Experiment 1 and 2 are utilized for a combined modeling approach in Chapter 7.

7 Modeling of take-over performance

Results from Chapters 5 and 6 represent the empirical basis for answers and conclusions in reference to changes in the driver state and potential consequences for the take-over performance. The chapters feature either a detailed analysis of results or the summary of findings depicted in the corresponding publication (Radlmayr, Fischer, & Bengler, 2019). Both experiments were conducted during the same period of time and in identical settings concerning the simulator. The experimenters took turns in accompanying participants for both Experiment 1 and 2. Based on recommendations from literature (Chapter 2) to include individual differences and the predisposition of drivers, the research question RQ5 was derived (Figure 7.1), aiming for a quantitative insight and comparison between all potential factors affecting take-over performance. Regarding the settings of Experiment 1 and 2, a joint modeling approach appeared to be a promising method. The modeling approach in this chapter is partly based on the approach from Gold (2016), for which the data from many experiments were aggregated using multiple linear and logistic regression. In addition, Gold (2016) concludes that a

> "[...] mixed-effect regression revealed a large variance induced by the drivers and significant model improvements under consideration of the drivers' predisposition. This contribution of the driver could and should be considered as a predictor, but is only available for drivers with several recorded take-overs or retrievable by driver monitoring."

The final model equations in Gold (2016) are focused on providing a reasonable prediction of TOT and quality aspects regardless of drivers' predisposition and current driver state. While the modeling presents a huge leap concerning the comprehensive understanding of take-over performance, the selection of empirical data from Experiment 1 and 2 was influenced by the conclusion from the modeling approach. More specifically, drivers' predisposition and measurement of the driver state were included in both experiments. In the modeling approach presented in this chapter, both predisposition measures and state changes as predictors were included to allow a better enlightenment of variance in the outcome variable. The mixed-model approach includes fixed and random effects. The selection of the models, the iterative process and final models and conclusions are detailed in this chapter. The modeling approach itself is based on tutorials from Singmann and Kellen (2017) and Winter (2013) using the packages *lme4* (Bates, Mächler, Bolker, & Walker, 2015), *afex* (Singmann, Bolker, Westfall, & Aust, 2019) and *ordinal* (Christensen, 2019) in *R* (R Core Team, 2018).

Driver State

Modeling of take-over performance with mixed effect models

RQ5 How do effects from e.g. driver state changes, situational factors and NDRTs on take-over performance compare to the effect of predisposition and individual differences of drivers?

Chapter 7

Figure 7.1: Research Question RQ5 that is addressed in the modeling approach.

7.1 Introduction, motivation and method

The take-over is of highest interest concerning the safety and comfort of CAD. Research and corresponding results in the last years have led to a broad understanding of various effects influencing the take-over process, see Chapter 2 and the empirical work in this thesis. A logical step is the integration of various effects into a model allowing the prediction of take-over performance. The example of Gold (2016) underlines the feasibility of predicting take-over performance. The integration of these models into real vehicles could lead way to the comprehensive understanding of all major effects giving way to a sound assessment whether - in fact - a driver is ready as fallback level prior to a take-over or not. The empirical data from Experiments 1 an 2 could serve for an additional modeling approach increasing the prediction quality by integrating information on the individual drivers and their state. This is advised by Gold (2016) in order to increase the amount of variance in the outcome parameters (measures of take-over performance) which is accounted for by the chosen model.

Based on the data from the two experiments in this work and the depicted advance in Gold (2016), the approach in this chapter is differing. Concerning the amount of data and the chosen predictors, a validation based on published results from other researchers does not seem feasible because the sensors on driver state are not commonly integrated in take-over research. While eye-tracking data can be found in some of the published work and could technically be utilized to validate the existing model, the seat pressure mats represent a new way of assessing activity prior to a take-over in experiments on CAD. Therefore, an alternative validation of the models in this work could be achieved by applying a data-split and using e.g. 70/30 percent of data to build and then validate the model. Regarding Figure 7.2 and the number of all potential predictors (19) in this

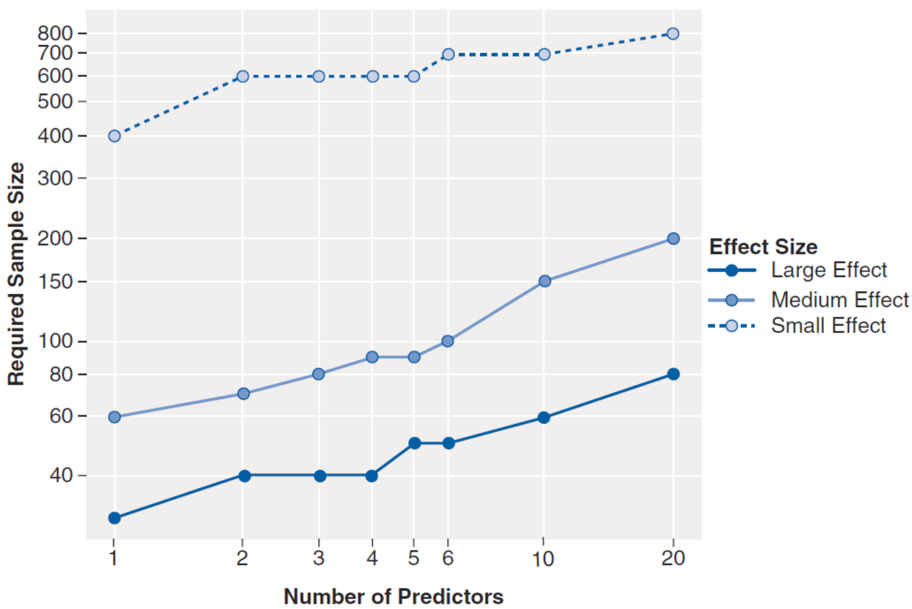


Figure 7.2: Rule of thumb on how big the sample size should be depending on the number of predictors and the expected effect sizes taken from Field et al. (2012).

work (see Table 7.2), with the current set of data points an identification of medium effects seems feasible. Reducing the available data points to allow a validation of the models by splitting the data would consequently decrease the number of identifiable medium effects and is not regarded in this modeling approach. In addition to the reflections on validating the models, the underlining focus of the modeling approach must be considered: while the resulting models from this chapter are not necessarily ideal for a future integration into real vehicles, considering the sensors used to assess the driver state, a comparison of hierarchical-build models allows a fundamental insight into whether the driver state and individual dispositions should be regarded in CAD or not. If non-mixed models allow fairly similar prediction qualities of take-over performance compared to models including these measures, future applications benefit from the quantitative comparison.

In this thesis, a step-wise modeling approach shedding light on the individual contribution of each predictor was chosen. In order to allow a understanding of the modeling approach presented here, the following section briefly introduces the group of models utilized.

Linear mixed effect models

A general introduction to multiple linear regression and the resulting fixed-effects models will not be provided here. Gold (2016) provide a thorough explanation of the main reasoning, most important key figures and their application in the field of predicting take-over performance. This chapter will provide a brief introduction to mixed effect models as applied in the following modeling approach. The introduction to linear mixed effect models here is partly based on the tutorial from Winter (2013) and relies strongly on the tutorial from Singmann and Kellen (2017) since the provided approach is focused on the field of experimental psychology. The example centers on linear mixed models, while the multinomial mixed logistic models utilized for prediction of categorical longitudinal and lateral accelerations follow the same reasoning.

The main reason for choosing mixed effect models in this approach are listed.

1. Both single and multiple regression depend - among other conditions - on the assumption that the underlying data are independent and the errors are identically distributed (Singmann & Kellen, 2017). This can be summarized to represent individual data points regarded for regression modeling to originate from only one participant. In case of repeated-measures designs this assumptions is violated since one participant provides more than one data point, in this case take-over. This holds true for both the data from Experiment 1 (three take-overs from one participant) and Experiment 2 (four take-overs from one participant). Consequently, linear multiple regression would not suffice to capture the underlying effects in a modeling approach whereas linear mixed effect models provide the necessary structure to allow the definition of additional random effects such as, e.g. the driver. While linear multiple regression provides good robustness towards violations of other assumptions, e.g. assumption of variance homogeneity, it is not robust concerning the assumption of independent data (Singmann & Kellen, 2017).

2. Linear, mixed effect models can handle missing or partly missing data entries, e.g. by applying partial pooling. Due to the tracking quality of the eye-tracker, some entries from Experiments 1 and 2 feature incomplete entries. While these entries could not be considered for linear, multiple regression, they are used in the following mixed effects' approach.

3. The increase in computational power in the last years has provided the resources needed to more easily compute linear mixed effect models and yielded the intro-

duction of respective packages in *R*, such as lme4 (Bates et al., 2015) or afex (Singmann et al., 2019).

Concerning a general equation of multiple regression, a multiple linear regression can be expressed as

$$\boldsymbol{y}_{n,i} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{x}_{1,i} + ... + \boldsymbol{\beta}_n \boldsymbol{x}_{n,i} + \boldsymbol{\epsilon}_{n,i} \tag{7.1}$$

In equation 7.1, the outcome y is modeled for the i^{th} participant, β_0 represents the grand mean or intercept, and β_1 to β_n represent the coefficients of the individual predictors x_1 through x_n. Utilizing known effects, a simple example is provided to allow a more comprehensive understanding of the final mixed effect models with regard to take-overs in CAD.

Suppose, the outcome y representing the TOT is being modeled using one predictor, traffic density (x and corresponding coefficient β_1). A final equation would yield β_0 and β_1 as results. β_0 representing the intercept would represent the value of TOTs in seconds that represents the mean of TOTs found in the underlying data set, e.g. 2.5 seconds. Depending on β_1, which in this example would correspond to the influence of traffic density and will be assumed to be 1, "one increase" in traffic density (e.g. from no traffic to 20 vehicles/km), the TOT would increase by 1 second, resulting in 3.5 seconds. The linear model provides a linear link between the outcome "TOT", the intercept and the influence of traffic density. The term ϵ is representing any form of variance observed in the data that cannot be explained by the linear link that was defined to consist of the intercept and the traffic density. This very basic example provides the basis for a linear **mixed** effect model representing a simple linear **fixed** effect model. TOT is modeled using the fixed effect traffic density.

Suppose, the data on which the model was build upon do not come from independent participants but every participant would experience four take-overs in total that were all considered in the model. Naturally, one participant could be generally faster than another participant a - currently not considered **random** - effect, that would increase variance in the outcome TOT, which could not be explained by the linear link. The difference between individual participants that is adding to the error term $\epsilon_{n,i}$ in the fixed effect model can be modeled using random intercepts. These random intercepts reduce the variance stemming from a difference in general "quickness of take-over" captured by the error term by adding a random effect, S_0. The equation including the random intercept would result to

$$\boldsymbol{y}_{n,i} = \boldsymbol{\beta}_0 + \boldsymbol{S}_{0,i} + \boldsymbol{\beta}_1 \boldsymbol{x}_{1,i} + ... + \boldsymbol{\beta}_n \boldsymbol{x}_{n,i} + \boldsymbol{\epsilon}_{n,i} \tag{7.2}$$

where, following the explanation from Singmann and Kellen (2017), the term $S_{0,i}$ "corresponds to the idiosyncratic effect associated to participant i". Concerning the example of TOTs being modeled by traffic density, the intercept or grand mean β_0 is now being adjusted by an additional, individual "grand mean" for every participant, representing their individual "quickness of take-over". The correction of the intercept β_0 by the random intercepts for every participant is based upon one participant contributing four take-overs to the data pool. In case the data points were independent, the definition of a random intercept per participant would not be possible.

Staying with the basic example, the influence of traffic density in the fixed effect model 7.1 has the same value for all participants. For every "increase" in traffic density, e.g. by one, the TOT would increase by one second (β_1 was identified to be 1 in this example) leading to a take-over time of 3.5 seconds. Similar to the random intercepts representing a difference in "general quickness", a mixed effect model also incorporates random

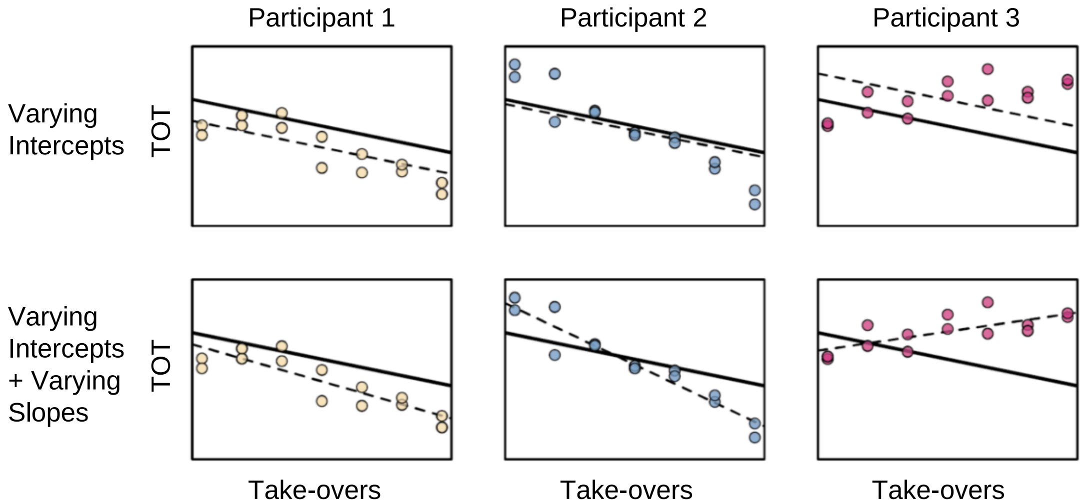


Figure 7.3: Examples of random intercepts and random slopes modified from Winter (2019).

slopes which represent individual differences with regard to the influence of traffic density. Corresponding to the example, some participants might react to an increase in traffic density more severe than other participants. In the fixed effect model, β_1 is the same for all participants, representing the slope of the line modeling the effect of traffic density on TOT. Random slopes allow the introduction of individually different slopes representing the individual "severity of TOT-changes with regard to different traffic densities". In case participants would actually react differently to the same change in traffic density, a random slope term would allow a quantification of this effect resulting in the following equation

$$\boldsymbol{y}_{n,i} = \boldsymbol{\beta}_0 + \boldsymbol{S}_{0,i} + (\boldsymbol{\beta}_1 + \boldsymbol{S}_{1,i})\boldsymbol{x}_{1,i} + ... + (\boldsymbol{\beta}_n + \boldsymbol{S}_{n,i})\boldsymbol{x}_{n,i} + \boldsymbol{\epsilon}_{n,i} \tag{7.3}$$

Notice, that for every fixed effect x_n, a corresponding random slope S_n can be defined. The resulting model in equation 7.3 now represents a mixed effect model including fixed effects, random intercepts for individual participants and random slopes for individual participants. Figure 7.3 provides an additional example with a total of 14 data points or take-overs per participant. The bold line in each of the sub-figures represents an estimate of the TOT from a linear regression model. In case a mixed model with varying intercepts is used, the dotted line in the sub-figures in the upper half represents the individual intercepts for Participants 1-3 or their individual "general quickness" of taking-over. Notice, that participant 3 is not speeding up but becoming slower. Adding varying slopes accounts for individual reactions to - in this case - the "effect of take-overs" in general. To account for individual reactions to an exposure of more than one take-over, varying slopes are introduced.

Concerning the pooling of data from Experiments 1 and 2, different structures of mixed effect models could be possible. Regarding the explanations from Singmann and Kellen (2017) and Winter (2013), different "items", in this work represented by take-over situations, could also account for random effects on the outcome. While the reasoning to incorporate these effects explicitly in the proposed mixed model is sound, they are not regarded in this approach: Different items accounting for variance in the outcome with regard to modeling take-over performance in CAD represent different take-over situations in which participants have to take-over. If participants would experience some of the take-overs in the construction site and some of the take-overs in the crash site, the manipulation of different situations presented more than once can be understood to incorporate a new, random effect caused by the specific situations picked from the population of all situations. While this effect is present in the data from Experiment 2 (Chapter 6), participants in Experiment 1 experienced a total of three situations which all differed: one crash site, one construction site and one interstate crossing. In order to quantify the random effect of situations in the mixed model, the data from Experiment 1 does not qualify, since the situations only appeared once and not twice as in Experiment 2. Following the reasoning of detectable effects and their corresponding effect sizes depending on the number of predictors and data points (Figure 7.2), the data from these two experiments need to be pooled together for a feasible modeling approach allowing a detection of medium effects. Consequently, the random effect structure (case (b) in Figure 7.4, items can be understood to represent situations) arising from different situations in the data pool from Experiment 2 is not regarded in this work. Case (c) in Figure 7.4 describes nested random effects, that represent "participants from different encroaching groups/experiments" (Singmann & Kellen, 2017). Since the pooled data in this work is originating from two different experiments, consideration of nested random effects could be feasible. However, both data sets were recorded during the same period of time, in the same simulator and both experimenters supported the other procedure and vice versa. The reasoning to suspect a nested random effect is highly unlikely and is not regarded in this work. Concerning the advice from Barr, Levy, Scheepers, and Tily (2013) to "keep it maximal" concerning the random effects structure, after thorough consideration of possible random effects, only the effect of several take-overs per participant is accounted for in the model. The resulting

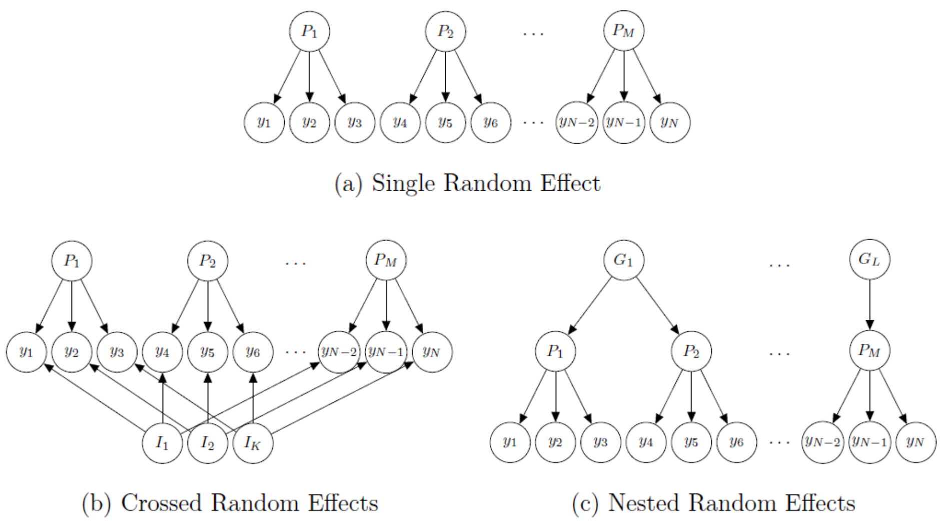


Figure 7.4: Examples of different structures of mixed effect models taken from Singmann and Kellen (2017).

modeling approach can be summarized to consist of mixed effect models incorporating a single random effect (participants, see case (a) in Figure 7.4). This single random effect incorporates random intercepts and random slopes for every participant and their potential correlation. Fixed effects are included in the modeling approach in an iterative process described in the following chapters.

Assumptions

Concerning the assumptions related to mixed models, the following list details the considerations of the approach conducted in this work.

- **Assumption of independence**
  A key assumption concerning multiple regression in general is the assumption of independent data, e.g. that all cases or data points considered in the modeling approach are independent. This assumption would be violated by the fact that participants experience more than one take-over. The resulting random effect is precisely the reason this modeling approach consists of mixed effect models. The models can deal with the violation of independence by quantifying the introduced random effect by providing random intercepts, slopes and their potential correlation.

- **Assumption of multicollinearity**
  Large degrees of multicollinearity should be avoided in order to meet the assumption of non-correlated predictors. For continuous predictors both various publications and guidelines on detection and thresholds (Mason & Perreault, 1991; Dormann et al., 2013) as well as potential "remedies" e.g. centering the data (Field et al., 2012) exist. In this approach, models are built and compared to one another. This incorporates testing for multicollinearity by providing the expected correlation of the regression coefficients of all predictor pairs (Baayen, 2008). Due to the bottom-up process of this modeling approach (detailed later in this section), predictors are dropped for the final modeling approach if their inherent contribution to the model shows non-significant results. While this approach is discussed critically in the discussion, it avoids a full check of all predictors concerning multicollinearity. To allow a comprehensive understanding of all potential associations between predictors, this section provides an analysis of multicollinearity between all predictors later on.

- **Assumption of homoscedasticity**
  A violation of this assumption indicating heteroscedasticity will be checked by assessing the residual plots of the fitted models. In case the plot gives way to suspect heteroscedasticity, results from the model have to be discussed critically concerning potential reasons for the violation.

- **Assumption of no/low effect of influential points**
  Typical ways of analyzing the effect from influential points in multiple regression consist of considering the DFBetas or the Cook's distance (Field et al., 2012). While these methods suffice for multiple regression, the calculation of these parameters for linear mixed effect models is considerably more complex. Nieuwenhuis, Te Grotenhuis, and Pelzer (2012) provide a solution based on built models from *lme4* that calculates both DFBeta-values and the Cook's distance to identify outliers and influential data points. The *influence()*-function from the corresponding package also allows an estimation of both parameters for the group levels defined by the introduction of single or nested random effects. A detailed description of the package and the calculation is provided in Nieuwenhuis et al. (2012) and serves as basis for the calculation of the Cook's distance in this work. The cutoff value of the Cook's distance is calculated by $4/n = .048$, n = number of group levels (e.g. participants in this work) following the rule of thumb from Nieuwenhuis et al. (2012). A plot is provided for identification of influential data points/participants and results are

regarded in the discussion of the final models rather than dropping outliers by default. In addition to the Cook's distance and the DFBeta-values, the *influence.ME*-package also provides the change in percent concerning the parameter estimation with/without all/the identified points:

> "For each higher level group, the percentage of change is calculated as the absolute difference between the parameter estimate both including and excluding the higher-level unit, divided by the parameter estimate of the complete model and multiplied by 100 %. A percentage of change is returned for each parameter separately, for each of the higher-level units under investigation" (Nieuwenhuis et al., 2012).

This value is referred to as "change with/without (in this work) participant" and will be added to analyze potential influential points.

- **Assumption of quantitative and categorical variable types**
  Concerning the definition of fixed effect predictors, linear mixed effect models in accordance with multiple regression rely on either unbound quantitative or categorical predictors with only two levels. In case a categorical predictor exceeds two levels (in this case e.g. type of NDRT), this issue needs to be solved using dummy coding, e.g. see the example of the rating of tasks regarding their manual and cognitive load proposed in Gold (2016). This is solved automatically in *R* when a categorical predictor with more than two levels is detected and is not addressed here.

- **Assumption of normally distributed residuals**
  Following the reasoning from Winter (2013) and Gelman and Hill (2006), multiple regression in general and linear mixed effect models accordingly show very high robustness towards non-normal distributions of residuals. Typically, this is analyzed by inspecting histograms of the residual distribution or Q-Q plots but is not be considered in this approach.

- **Linearity of the outcome, e.g. modeling a linear relationship**
  The premise of using linear mixed effect models always constitutes a linear relationship between the outcome and potential predictors. In case reasonable doubt exists that the relationship shows a non-linear nature, additional methods of modeling can be used, such as logistic modeling or generalized mixed models. While an introduction to the latter is not given here, the modeling of crash probability and longitudinal acceleration in Gold (2016) suggests that a linear mixed model is not feasible for both these outcomes in this work. The crash probability was modeled using a binomial mixed modeling approach, while the longitudinal and lateral acceleration were modeled using multinomial mixed linear regression modeling.

Sample, predictors and outcome variables

Multiple regression relies on complete data sets to allow a modeling approach. In case mixed models are utilized, non-complete data can be regarded for the modeling, increasing the amount of total information available for the modeling.

The data set for this modeling approach consists of data from Experiment 1 and 2. A total of 299 data entries were regarded for the modeling approach, with 243 cases consisting of complete entries. Main reason for incomplete data entries can be identified by considering the predictors used for the final models. While trait-measures, such as sex, age or driving experience were available for all participants due to demographic questionnaires, the inclusion of state measures from eye-tracking and seat pressure mats introduced the number of incomplete entries. Seat pressure mats proved to be a reliable source of driver activity in the seat with almost no issues of data loss during the experiment. The eye-tracking data was cut to periods of one minute or 10 seconds to allow an assessment of both absolute values in the last minute/10 seconds prior to the take-over as well as changes manifesting during the period of automated driving. Before the final output, the eye-tracking data was checked for freezes or data inconsistencies such as values of below 70 % for the reported detection rate. In case freezes or data inconsistencies were detected, the affected period of one minute/10 seconds was disregarded for analysis.

The sample of 299 entries consisted of 84 participants, with 33 females and 51 males, with a combined mean age of 31.41 years (SD = 13.58 years), and a range of 17[1] to 73 years of age.

A list of outcome variables is provided in Table 7.1 including the potential range of values and the actual range of the data to allow an assessment of boundary issues. In addition, the predictors from the final model equations without the actual equation from Gold (2016) are provided including the predictors that are regarded in this work. In case predictors from Gold (2016) were disregarded, the underlying reasoning is provided in the following list.

Table 7.1: List of outcome variables considered in this modeling approach including their potential range of values and the actual range found in the data. In addition, any predictors from the final model equations from Gold (2016), that were also considered in this modeling approach, are provided.

Name	Potential range	Actual range	Considered predictors in (Gold, 2016)	Considered predictors in this modeling approach
Take-over time (TOT) [s]	0 – 7	.60 – 6.18	Time-budget, lane, traffic density, repetition, age	traffic density, age
Lateral acceleration [m/s^2]	0 – 10, unilateral	.15 – 8.42	Time-budget, lane, traffic density, repetition, age	traffic density, age
Minimal longitudinal acceleration [m/s^2]	-10 – 0	-10.23 – 0	Time-budget, traffic density, load	traffic density, NDRT (representing load)
Time to collision (TTC) [s]	0 – 7	.19 – 6.22	Time-budget, lane, traffic density, repetition, load, age	traffic density, NDRT (representing load), age
Probability of crash [ ]	0, 1	0, 1	Lane, traffic density, repetition	traffic density

[1] Accompanied driving is legal by the age of 17 in Germany. In this case of participating in the experiment in a driving simulator, no legal guardian was present.

- **Gaze reaction time**
  In this modeling approach, the gaze reaction time is not considered: Only some participants engaged in a visually demanding NDRT. In case participants did not engage in NDRTs or their visual attention was on the road regardless of the NDRT, no gaze reaction time is available. In addition, following the modeling approach by Gold (2016), gaze reaction times were best accounted for by the intercept. This represents participants reacting quickly to the salient stimulus of the RtI without incorporating more complex cognitive processes in comparison to the TOT.

- **Take-over time (TOT)**
  The time budget was set to seven seconds throughout both experiments, rendering the predictor not feasible since no variance was introduced. The same reasoning applies to lane. All participants were on the middle lane when the RtI was issued. Repetition was not considered since the model consisted of a mixed model approach, incorporating more than one take-over per participant through considering the random effect participant.

- **Probability of brake reaction**
  The prediction quality significantly improved by splitting between drivers which reacted by either braking or steering in Gold (2016). While the feasibility of this approach is observed, following Figure 7.2, the data set did not allow splitting the data and this outcome was not regarded in this work.

- **Lateral acceleration**
  Analogue to the reasoning for TOT, the predictors time budget and lane were not considered as predictors for the maximal lateral acceleration. The predictor repetition was accounted for by the random effect of the linear model.

- **Minimal longitudinal acceleration**
  The final modeling approach from Gold (2016) showed that the prediction quality increased when the longitudinal acceleration is not modeled on a continuous interval scale but the probability of either an absolute small [<3.5 m/s^2], medium [3.5 m/s^2 – 7.0 m/s^2] or large [>7.0 m/s^2] acceleration is modeled. The modeling approach in this work started with a linear mixed effect model, since the outcome variable does not necessarily has to be normally distributed (Field et al., 2012). In addition, the distinction in participants reacting by braking or steering could be nested within them and is accounted for by the random structure of the mixed model approach. Concerning the predictor load, the NDRTs and no engagement in NDRTs were represented by "load" in Gold (2016). In this work, the NDRTs and no engagement are modeled as factorial predictor since *R* uses dummy modeling in the background to account for factorial predictors with more than two levels. Time-budget was not accounted for since all take-overs featured a time budget of seven seconds.

- **Time to collision (TTC)**
  The TTC in this work is modeled including traffic density, load and age, excluding the other predictors as explained above. Concerning the range, the theoretical value of zero represents a crash with the obstacle for the crash situation. The crash probability is modeled separately, so TTC values of zero are excluded from the analysis of the TTC for all situations. In case participants came to a full stop after the RtI before the system limit, the TTC would reach infinite values before the calculation

rendered void results due to the division by zero (TTC = distance to system limit divided by current velocity). While these values are not feasible since the minimal TTC is of interest, the division by zero was accounted for in the processing of data. For velocities of <1 km/h the TTC is not calculated and could not be regarded for minimal TTC for this modeling approach.

- **Probability of a crash**
  The crash probability was modeled in accordance to Gold (2016) using the traffic density. Lane and repetition were not regarded due to the explanation above.

The modeling approach in this work is focused on comparing various predictors concerning their contribution to allow a prediction of the outcome measures provided in Table 7.1. The predictors used in the modeling approach from Gold (2016) serve as starting reference for the approach in this work and can be found in the far right column of Table 7.1.

The main addition in this approach consists of inclusion of additional predictors, such as participants' traits (driving experience or annual mileage) and driver state changes measured using eye-tracking and seat pressure mats. These include e.g. PEOR or changes of the COP from the seat pressure mats. Table 7.2 provides an overview of the predictors that are regarded in the modeling approach in this work. Driver state changes are also assessed in % between the second and the last minute before the RtI. The second minute (and not the first) is utilized to allow an identical length of the considered interval and to avoid including "settling into automated driving mode"-effects. Data were cut moving backwards from the point of time of the RtI.

Table 7.2: List of predictor variables initially considered in this modeling approach including their range found in the data.

Name	Range	Assessed interval
	Random effects	
Participant/Nr. [ ]	3 or 4 take-overs	To account for multiple take-overs per participant, the random effect "nr" is introduced.
	Fixed effects	
	Traits	
Sex [ ]	male or female	No other sexes were provided by participants, the factor is considered to be exhaustive.
Age [years]	17 – 73	Modeled both linear and as 2nd-degree polynom, regarding results from (Gold, 2016).
Possession of valid driver's license [years]	0 – 56	Dropped for the final approach due to a strong correlation with age.
Kilometers per year [km/year]	Categorical: <5k, 5k – 10k, 10k – 20k, >20k	The levels represent small, medium, large and very large mileage per year.
Subjective driving style [ ]	Likert-scale: -2 – 2	-2 represents a subjectively defensive driving style, 2 represents a sporty driving style.

	Situational factors	
Traffic density [vehicles/km]	Categorical: 0 or 20	Traffic density during CAD and the take-over situation.
Situation [ ]	Categorical: crash site, construction site, interstate crossing	To account for more variance in the outcome[2], apart from traffic density, situation is also considered.
	State changes	
Type of NDRT [ ]	Categorical: no, SuRT, n-back, motoric	While some participants were instructed to engage in NDRTs and others could choose freely, the type of NDRT represents the engagement in a task right before the RtI was issued. Duration of engagement prior to the RtI is not considered as predictor since most intervals lasted no longer than one minute, both for instructed and free engagement.
HGD (SD of horiz. gaze position) [m]	.02 – .3	The value per participant represent the mean HGD during the last minute prior to the RtI as measure of the horizontal tracking activity.
Changes of the HGD [%]	-90.82 – 166.59	Changes of the driver state were evaluated by assessing the difference in % of the HGD between the second minute of automated driving and the last minute before the RtI.
PEOR [%]	.07 – 99.90	Values represent the total PEOR during the last minute before the RtI.
Changes of PEOR [%]	-99.90 – 356.27	The difference in PEOR between the second and last minute before the RtI.
PEOR(10s) [%]	0 – 100	The values represent the PEOR in the 10 second interval before the RtI to allow a better understanding of visual attention.
Blink duration [s]	.10 – .62	The mean blink duration during the last minute before the RtI.
Changes of blink duration [%]	-63.75 – 101.72	The difference in blink duration between the second and last minute of automated driving prior to the RtI.
Blink frequency [Hz]	.02 – 1.42	The mean blink frequency during the last minute before the RtI.
Changes of the blink frequency [%]	-100.00 – 400.00	Changes in blink frequency between the second and last minute before the RtI.

[2]Gold (2016) revealed significant influences from situational predictors, such as time budget or lane. In this work, these were not varied, so they cannot be used as predictors. To allow a more comprehensive understanding of both situational and driver state predictors, the categorical predictor situation is introduced.

Changes in the center of pressure (COP) for the seat [%]	-56.55 – 80.64	The mean difference of the COP in the seat between the second and last minute of automated driving before the RtI. This represents an in/decrease in activity in the seat.
Changes in COP for the backrest [%]	-33.35 – 116.55	The mean difference of the COP in the backrest between the second and last minute of automated driving before the RtI. This represents an in/decrease in activity in the backrest accordingly.

Prior to the detailed depiction of the modeling approach in Section 7.1 including a check for multicollinearity for the final model, all potential predictors listed in Table 7.2 were checked for potential associations. While continuous/numeric predictors can be tested for multicollinearity using standard procedures like Pearson's or Spearman's correlation or evaluating the variance inflation factor (VIF, (Montgomery & Peck, 1992)), these options are not available for categorical predictors. The final model uses dummy coding to allow an appraisal of multicollinearity which is not sufficient for a detailed understanding of the various associations between predictors.

Since the list of predictors also features categorical ones such as situation or type of NDRT, these associations should be assessed. A total of three combinations is possible: continuous vs. continuous, continuous vs. categorical and categorical vs. categorical. The following list addresses the method of checking for either a correlation or general association between all possible combinations of predictors and is based on a procedure from Brandl (2019).

- **Continuous vs. continuous predictor**
  A Spearman's correlation is calculated to determine if multicollinearity can be detected. Results greater than 0.7 are discussed critically following a general rule of thumb (Mason & Perreault, 1991).

- **Categorical vs. categorical predictor**
  Two nominal predictors can be evaluated concerning their association using Cramer's V (Cramér, 1999) which is based on Pearson's chi-squared statistic and allows an assessment of association between two nominal predictors. The result from Cramer's V test can be understood as the effect size for a chi-square test of association and is bias corrected to allow a comparison to the other association results (Mangiafico, 2016). The threshold is also set to 0.7 based on the correlations.

- **Continuous vs. categorical predictor**
  The relationship between any categorical and continuous variable can be assessed utilizing a one-way ANOVA. While the ANOVA only yields results on significant differences between levels of the categorical variable in the continuous one, it provides a fit for the individual group means. Comparing the fitted values with the observed data (in this case the continuous predictor), eta can be derived, providing a measure of association between the categorical and the continuous predictor (*Correlation between a nominal (IV) and a continuous (DV) variable*, 2014). The threshold is also set to 0.7 to allow a comparison based on an identical threshold between all three associations.

The full results can be found in Table B.1. A very high correlation was found between the predictors age and years of possessing a driver's license. This can be attributed to the fact that almost all participants received their driver's license at the age of 18, accounting for the correlation. Based on the results, it can be assumed that age represents both years living in addition to holding a driver's license. Thus, only age was kept as predictor for the modeling approach, while the other one was dropped.

In addition, type of NDRT showed a strong association with PEOR, both for 10 seconds and the last minute prior to the RtI. Since the NDRTs were either visual, cognitive, motoric or none at all, they are linked to the visual attention before the RtI. While the predictors appear to capture similar information, PEOR only links to visual attention, while the type of NDRT includes potential information if participants might have been looking but "failed to see" due to a high involvement in the cognitive or motoric NDRT. The categorical predictor NDRT also incorporates a limited combination with other predictors since Experiment 1 featured no NDRTs. Insufficient permutation of predictors can lead to strong associations (Gold, 2016) accounting for the results.

Following newer insights into multicollinearity and potentially related problems (Vanhove, 2019), correlated or associated predictors do not necessarily have to be dropped or substituted. Regarding statistical consequences, multicollinearity affects the estimates of regression coefficients. Coefficients that are influenced tend to vary more from sample to sample than estimates of regression coefficients that are not affected by multicollinearity (Vanhove, 2019). Regardless, multicollinearity does not bias the coefficient estimates since on average, the estimated coefficients equal the parameter's true value (Vanhove, 2019). Taking these findings into account, individual contributions of associated predictors might not yield significant results, but the prior identification helps in understanding and discussing results and is regarded in the discussion of the modeling approach.

Regarding the underlying research question of the modeling approach which is not aimed at providing model coefficients ready to be implemented but rather a comparison of potential effects, both PEOR predictors as well as type of NDRT were assessed.

A high correlation was also revealed between the PEOR in the last minute and the last 10 seconds together with a high correlation between the PEOR in the last minute and the change of the PEOR. The correlation between the absolute values of the PEOR can be accounted for by participants most likely not changing their behavior frequently within the last minute before the RtI. In case participants were engaged in the visual NDRT, this was likely true for both 10 seconds and the last minute before the take-over. In addition with all other conditions (no NDRT, cognitive, motoric) providing no visual stimuli except the environment outside of the vehicle and thus a limited introduction of variance in the PEOR values, the correlation is regarded to be non-critical concerning the modeling approach. Following the same reasoning based on Vanhove (2019), the insight into which duration of PEOR could be sufficient for effects in take-over performance, all three predictors were regarded for the modeling approach.

Procedure

This chapter provides an overview on how the modeling approach in *R* is conducted.

Concerning the random-effects-structure of the model, the "maximal" model should be considered starting any hierarchical approach (Barr et al., 2013). The *mixed*-function of *afex* is built upon the *lmer*-function building a linear mixed effect model incorporating the fixed and random effects specified. In addition, it allows the comparisons of the contribution of both fixed and random factors by comparing the underlying models with and without all fixed and random factors. If the *lmer*-function of *lme4* was used to directly fit a model, Singmann and Kellen (2017) point out that this approach

> "[...] is associated with the problems already noted above. First, lmer does not provide p-values so that one needs to perform an additional inferential step. Second, the default contrast codes in R are such that model with categorical covariates (i.e., factors) produce parameter estimates that do not accurately represent lower-order effects (e.g., main effects) if higher-order effects (i.e., interactions) are present. This latter fact is the reason that some people recommend to transform factors into numerical covariates by hand. R contains coding schemes that are orthogonal and do not have this problem. The easiest way to change the coding globally is via the afex function set-sum-contrasts."

In addition, orthogonal sum-to-zero contrasts are often a more reasonable default than treatment contrasts for experimental designs (Singmann & Kellen, 2017). The *afex*-package provides the possibility to assess p-values by using orthogonal sum-to-zero contrasts in the default setting and calculates them for the terms/factors in the mixed model using the following methods (if not specified, the "Kenward-Roger-method" is used as default): Kenward-Roger, Satterthwaite, Likelihood Ratio Test (LRT) or parametric bootstrap (Singmann & Kellen, 2017). In this work, the models are always fitted using a restricted maximum likelihood estimation (REML), since the Kenward-Roger approximation requires the model to be fitted with REML (Singmann & Kellen, 2017). For more information on a detailed description of the process refer to page 29 of Singmann and Kellen (2017). Thus, a "maximal" model in this work would incorporate all fixed and random effects specified in Table 7.2. The *mixed*-function of *afex* would then

> "fit[s] an encompassing model with all parameters and one reduced model corresponding to each of the model terms in which the parameters corresponding to the term are withhold from the full model (all fits are performed with lmer()). [...] After estimating all necessary models the p-values are calculated with the corresponding method" (Singmann & Kellen, 2017).

However, this approach yielded various errors and convergence problems of the underlying algorithm when the fist maximal model for TOT was defined in this work. While Singmann and Kellen (2017) provide possible explanations and solutions to this problem, the approach in this work deviated due to the random-effects-structure: Table 7.2 provides a list of all 18 fixed effects for which - technically - the maximal model would account for the accompanying random effects, including random intercepts, random slopes and their corresponding correlation for each of the 18 fixed factors. Including more than one fixed effect into the maximal model and specifying the accompanying random effects would therefore be impossible to analyze if convergence problems were detected. Consequently, the approach in this work consisted of a "bottom-up"-approach, where the maximal model was built only including one fixed effect including the maximal random effects structure for this fixed effect (intercepts, slopes and their correlation).

In case convergence problems were encountered, the random-effects-structure is reduced by first ignoring a potential correlation between random intercepts and random slopes. If convergence problems persisted, random slopes were ignored. In case convergence problems were still reported, the underlying data structure does not benefit from specifying random effects and the random effect for this fixed factor was ignored completely. By defining only one fixed effect and the accompanying random structure for this fixed effect, the *mixed*-function naturally provided the p-value for this fixed effect in comparison to the model including only the grand intercept. In case the p-value did not reveal significant results for the said fixed effect, it was not included for further modeling building. In addition, the Akaike information criteria (AIC) was always assessed to compare the specific models since R^2 always rises by including more (fixed) predictors while the AIC penalizes additional predictors. In case the AIC would decrease, the model fit was assessed to be better (Akaike, 1974). The Pseudo-R^2 calculated for individual models represents the amount of variance explained by the model fit and would be 1 for a perfect fit. In addition, the root mean square error (RMSE) is reported to allow an understanding of the fit in units of the outcome.

The following example for TOT illustrates the approach from this work and is taken from the analysis found later. To illustrate the specific process in *R*, the *R*-specific formula annotation will be used.

1. TOT is modeled starting with the first fixed effect "sex" and the accompanying full random effect structure including random intercepts, random slopes and their correlation for all take-overs from one participant, identified by his/her number "nr":

   `tot ~ sex + (sex|nr)`

   Note: `(sex|nr)` expands to `(1 + sex|nr)`.

2. Convergence problems were reported, indicating that the random-effects structure is too complex for the actual underlying data.

3. First step of reducing the random-effect structure: ignoring the correlation between random intercepts and random slopes:

   `tot ~ sex + (sex||nr)`

   Note: `(sex||nr)` expands to `((0 + sex|nr) + (1|nr))` and represents random intercepts and random slopes for participants (nr), but not the correlation between them.

4. Results still yielded convergence problems, and the random-effects structure was reduced to only random intercepts:

   `tot ~ sex + (1|nr)`

5. Results showed no issues concerning convergence and the model was identified to represent the maximal random structure concerning the fixed effect "sex".

6. Comparing the fixed effects, in this case the influence of "sex" on TOT in comparison to modeling TOT by just the grand intercept (of all participants), the p-value of .82 (mixed-model-ANOVA-table: $F(1, 80.65) = .05$, type-3-tests, Kenward-Roger-method) suggests, that "sex" explains a non-significant amount of variance in TOT.

7. Any additional models for TOT will not incorporate the fixed factor "sex".

8. This process was carried out for all remaining fixed effects and their accompanying random effects structure. In case a fixed effect showed significant results, it was incorporated in the following models.

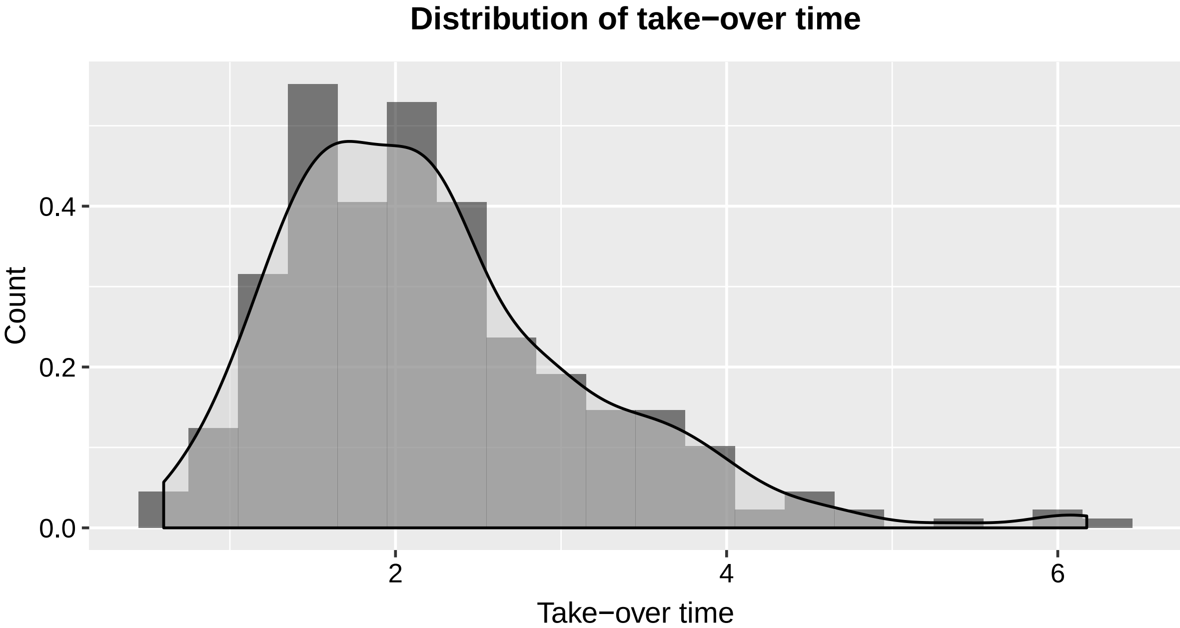


Figure 7.5: Histogram and density plot of the take-over time distribution. The linear mixed modeling approach incorporating random effects was based on visual examination of this plot.

7.2 Results

7.2.1 Modeling of take-over time

TOT was modeled as described in the process aforementioned. The distribution of the data set was plotted in Figure 7.5.

Table 7.3: Hierarchical approach for TOT. In case only random intercepts were supported, random slopes and the correlation between random slopes and random intercepts showed convergence problems. If only random intercepts and random slopes were supported, their correlation showed convergence issues.

Fixed effect	Maximal random-effect structure supported by the data (no convergence problems)	Contribution of fixed effect (compared to model without it, Kenward-Roger-method)
	Traits	
Sex	Random intercepts: `(1|nr)`	$F(1, 80.65) = .05, p = .82$
Age (linear)	Random intercepts: `(1|nr)`	$F(1, 80.11) = 8.34, p = .005$
For the following models, "age" (linear) is always considered as fixed effect, but is not reported again for the individual model comparisons which include models with/without "age" also.		
Mileage per year	Random intercepts: `(1|nr)`	$F(1, 78.14) = .08, p = .97$
Subjective driving style	Random intercepts: `(1|nr)`	$F(1, 78.76) = .19, p = .66$
	Situational factors	
Traffic density	Random intercepts: `(1|nr)`	$F(1, 97.29) = 1.88, p = .17$

Situation	Random intercepts: `(1|nr)`	$F(2, 240.50) = 25.49$, $p < .001$
For the following models, situation is always considered as fixed effect, but is not reported again for the individual model comparisons which include models with/without situation also.		
	State changes	
Type of NDRT	Random intercepts: `(1|nr)`	$F(3, 267.71) = 1.23, p = .30$
HGD (last minute prior to RtI), (n = 287)	Random intercepts and slopes: `(hgd(lastmin)||nr)`	$F(1, 186.69) = .04, p = .83$
Change of the HGD, (n = 274)	Random intercepts, slopes and their correlation: `(hgd(lastmin2 perc)|nr)`	$F(1, 43.26) = .04, p = .84$
PEOR (last minute), (n = 282)	Random intercepts: `(1|nr)`	$F(1, 270.94) = 2.88, p = .09$
Change of the PEOR, (n = 269)	Random intercepts: `(1|nr)`	$F(1, 253.09) = .26, p = .61$
PEOR (10 seconds)	Random intercepts: `(1|nr)`	$F(1, 290.45) = 4.58, p = .03$
For the following models, PEOR(10s before RtI) is always considered as fixed effect, but is not reported again for the individual model comparisons which include models with/without PEOR(10s before RtI) also.		
Blink duration, (n = 263)	Random intercepts: `(1|nr)`	$F(1, 256.93) = .19, p = .66$
Change of the blink duration, (n = 250)	Random intercepts: `(1|nr)`	$F(1, 226.86) = .18, p = .67$
Blink frequency, (n = 263)	Random intercepts, random slopes and their correlation: `(blink frequency(lastmin)|nr)`	$F(1, 32.83) = 1.38, p = .25$
Change of the blink frequency, (n = 251)	Random intercepts: `(1|nr)`	$F(1, 244.43) = .01, p = .94$
Change of the COP in the seat	Random intercepts and random slopes: `(copseat(changes)||nr)`	$F(1, 46.12) = 1.41, p = .24$
Change of the COP in the backrest	Random intercepts and random slopes: `(copback(changes)||nr)`	$F(1, 36.88) = .02, p = .90$

The following paragraph focuses on the significant results from the modeling approach for TOT.

Age shows significant results, modeled with a linear term. The underlying data for Age only supports random intercepts without showing convergence problems when fitting the model. The predictor Situation shows a highly significant contribution to the model, in line with individual results from Experiments 1 and 2. The data structure supports random intercepts for the factor Situation. While the predictor PEOR during the last minute before a RtI shows a tendency for significant findings, the PEOR in the last 10 seconds reveals a significant effect. In addition, the individual data from participants supports random

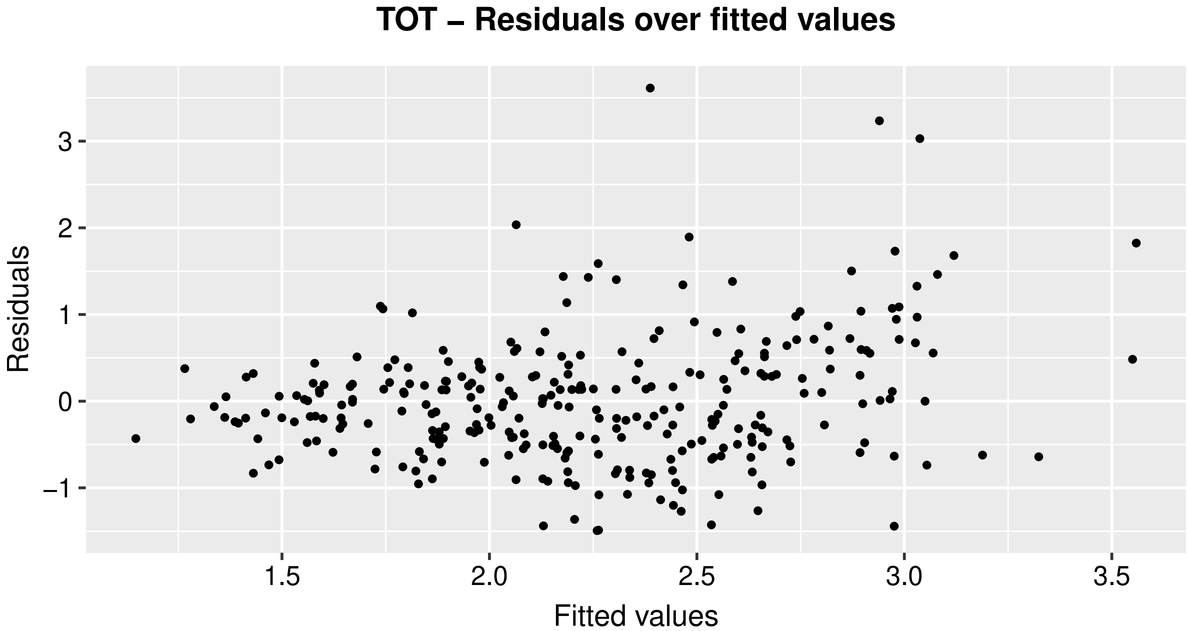


Figure 7.6: Plot of the residuals over the fitted values of the final model for the TOT. The distribution is not suggesting heteroscedasticity with few potential outliers. The assumption of homoscedasticity is not violated following this plot.

intercepts, translating to participants "starting" at different individual levels of "take-over quickness" depending on their visual attention right before a RtI.

Based on the hierarchical approach depicted in Table 7.3, which shows the process from top to bottom, the final model for TOT was derived. Concerning the assumptions, the residuals plotted over the fitted values in Figure 7.6 do not show clear signs of heteroscedasticity. Concerning the adjusted Cooks' distances plotted in Figure B.1, four participants (id nr. 19, 25, 30 and 40) are identified as influential. Visual analysis of the DFBeta-values and the changes in percent with/without specific participants is conducted using the plots in Figures B.2 and B.3. The DFBeta-cutoff value is set to $2/\sqrt{n}$ with n equals the number of higher level units following Nieuwenhuis et al. (2012) and David, Kuh, and Welsch (1980). For the TOT model, the DFBeta-value results in $2/\sqrt{84} = .22$. Analyzing the plots in Figure B.2, the four influential participants nr. 19, 25, 30 and 40 show DFBeta-values above the threshold. To identify the final model fit, an identical model excluding participant nr. 19, 25, 30 and 40 is fitted to allow a comparison. The final model description can be found in Table 7.4. In addition to the model description, the correlation matrix of the fixed effect is also reported in Table 7.5. Results show, that there is no issue with multicollinearity between the fixed effects as discussed in the list of assumptions. Age shows a strong correlation with the intercept. Plotting the final results of a mixed model approach cannot be done as straight forward as plots from linear regression since the random effects can introduce random intercepts and slopes. The individual fixed effects were plotted over the actual observations, including the 95th-confidence interval and the distribution visualized as rug-plot on the axis of the plots in case of continuous predictors. For the fitted model of TOT, plots result in showing the individual contribution of age, the situations and PEOR(10s) in Figures 7.7, 7.8 and 7.9. The plots are based on the fitted model including all participants. The reduced model was fitted to allow a

Table 7.4: Model description of final model for TOT. The values for the fitted model without the influential points are reported in brackets. The model was fitted on 296 (283) observations.

		Model fit			
AIC	Pseudo-R^2 (fixed effects)		Pseudo-R^2 (total)		RMSE
775.48 (677.94)	R^2 = .17, (.19) (fixed effects)		R^2 = .32, (.41) (total)		.73 (.62)
		Fixed effects			
Factor	Estimate	Std. error	t-value	df	p-value
Intercept	3.18, (3.18)	.18, (.18)	17.31, (17.56)	121.02, (111.59)	<.001, (<.001)
Age	-.01, (-.01)	.00, (.00)	-2.91, (-3.23)	77.31, (73.64)	<.001, (<.001)
Crash site	-.59, (-.58)	.10, (.09)	-5.98, (-6.78)	208.35, (199.10)	<.001, (<.001)
Interstate crossing	-.78, (-.65)	.16, (.14)	-4.89, (-4.51)	255.83, (238.95)	<.001, (<.001)
PEOR (10s)	-.003, (-.30)	.14, (.13)	-2.16, (-2.34)	290.42, (276.46)	.03, (.02)
		Random effects			
Group	Parameter		# Groups	Std. Deviation	ICC
Participant nr.	Intercept		84 (80)	.37, (.41)	.18, (.26)
Residual	–		–	.79, (.68)	–

Table 7.5: Correlation matrix of fixed effects. Note, that for linear mixed effect models, this matrix is "an approximate correlation of the estimator of the fixed effects" Baayen (2009). The significant fixed predictors in combination with results from the initial check on multicollinearity are deemed non-correlated whereas age shows a correlation with the intercept.

	Intercept	Age	Sit. crash site	Sit. int. crossing
Age	-.80			
Sit. crash site	-.22	-.01		
Sit. int. crossing	-.06	-.05	.32	
PEOR(10s)	-.46	.06	-.09	-.19

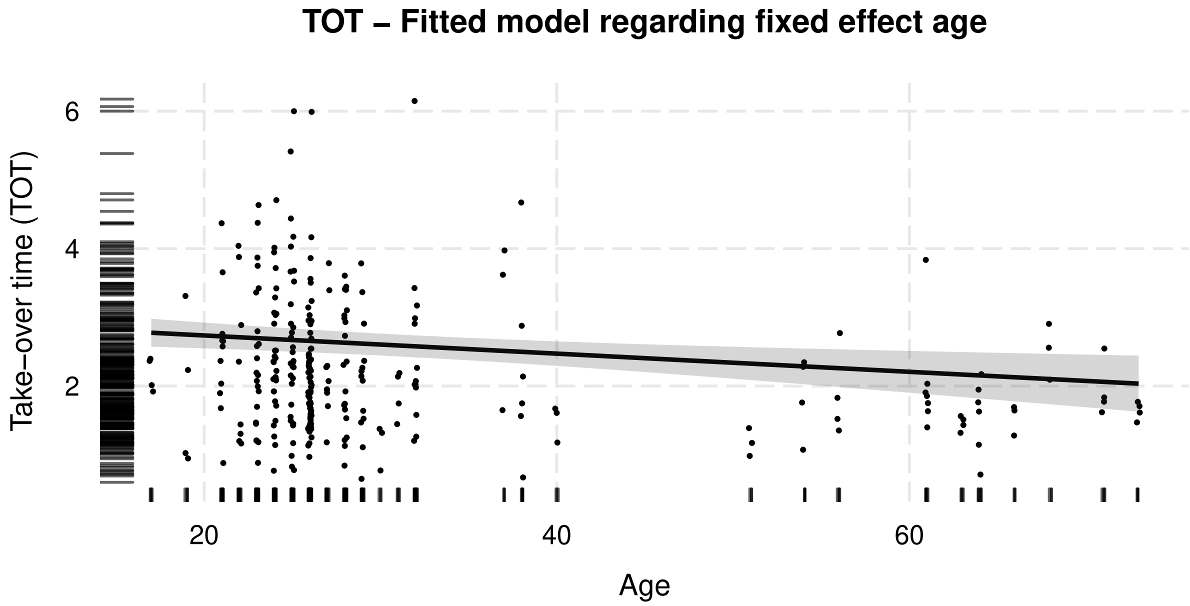


Figure 7.7: Plot of the fitted model for the fixed effect age only. The original data are also plotted.

better discussion of results. The final model plot in Figure 7.10 shows the fitted models for individual participants and represents a visualization of the random intercepts.

The hierarchical modeling approach revealed that for TOT and this data set, the global intercept is a good predictor. The fixed factor Age revealed a significant but small influence on TOT in negative direction. Older participants show reduced TOTs, but to a minor degree. The categorical factor Situation shows highly significant results. In Table 7.4, the construction site served as basis since it is alphabetically the first level. Compared to the construction site, both the crash site and the interstate crossing lead to significantly lower TOTs. In addition, the PEOR in the last 10 seconds before the RtI shows significant results. Participants with a higher amount of PEOR show faster TOTs. Comparing the extreme values of 0 PEOR or no visual attention on the road to 100 PEOR, participants would speed up .3 seconds. Concerning the random intercepts supported by the data and identified by the model, comparison of the standard deviations in Table 7.4 show that the amount of unexplained variance in the data ("residual") is still twice as high as the amount of variance explained by the random intercepts. In addition to the intraclass correlation coefficient[3] (ICC) and the comparison of calculated pseudo-R^2 between fixed effects only and total (Table 7.4), the introduction of random effects significantly improves the prediction quality. Regardless, the final fitted linear mixed model remains unable to predict TOT sufficiently by only accounting for a third of variance explained by the model (R^2(total) = .32). Results in brackets represent the model without the outliers and show an improvement, especially for the random effects. Checking the individual contribution of the outliers Nr. 19, 25, 30 and 40 both in Figure 7.10 and the raw data, Nr. 25, 30, 40 show one take-over with a TOT of approximately six seconds each, most likely accounting for the critical Cook's distance. Participant 19 revealed the lowest TOT with no visual attention in the interstate crossing, leading to the identification as outlier. Overall, outliers should

[3]Measure to show how much of the overall variance in the outcome (TOT) is accounted for by introduction of the random effects. If the ICC is close to zero, a simpler model without random effects should be fitted.

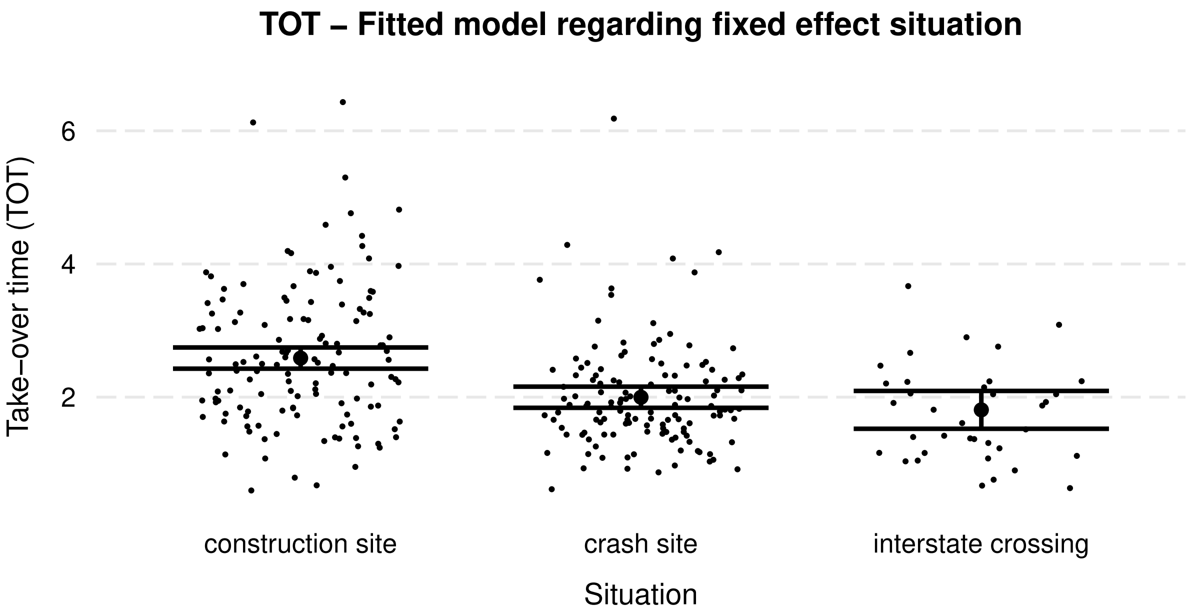


Figure 7.8: Plot of the fitted model for the fixed effect situation only. The original data are also plotted.

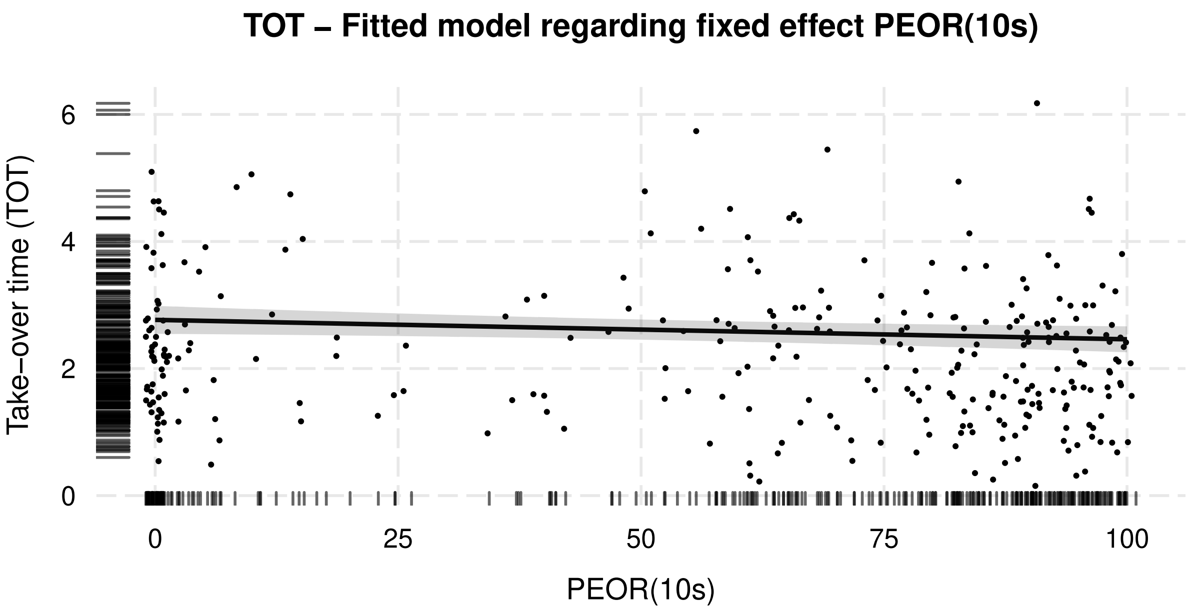


Figure 7.9: Plot of the fitted model for the fixed effect PEOR(10s) only. The original data are also plotted. Data are slightly jittered to allow better plotting, interval is 0 – 100.

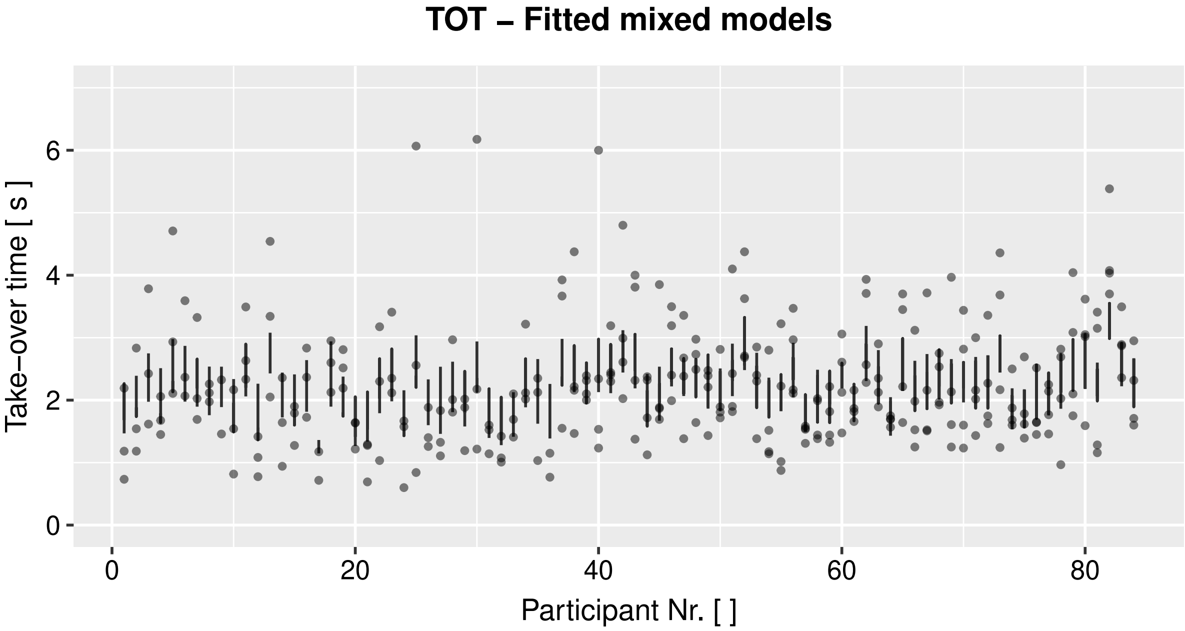


Figure 7.10: Plot of the fitted mixed models for the individual participants.

not be excluded from the modeling approach by default but analyzed in more detail to gain better insight into why these values were observed. In this case, no additional reasoning seems feasible to account for the extreme values of TOT.

A thorough discussion of the results and the limitations can be found in Section 7.3.

7.2.2 Modeling of crashes

Overall, nine crashes were recorded in the data set. Since the outcome crash can either occur or not occur, a linear mixed model approach including a continuous outcome must not be utilized. Following the example from Gold (2016), crash was modeled incorporating its binomial distribution. Typically, the approach would consist of a logistic regression which is a member of the class of generalized linear models (GLMs), but cannot incorporate mixed effects. As detailed in the introduction to this chapter, multiple take-overs from one participant violate the assumption of independence in addition with the research questions addressing the contribution of individual participants. Therefore, crash was modeled using generalized linear mixed models (GLMMs) that are able to incorporate the binomial distribution and a random effects structure. These models are also incorporated in the *afex*-package, are built using the *lme4*-package and the procedure is identical to linear mixed models. In order to determine the maximal random structure supported by the underlying data and to avoid a complicated search in case convergence issues became obvious in a model comparison with all fixed effects, the final GLMMs were built in the same hierarchical process as described for TOT. Fixed effects were added step by step to determine the maximal random structure supported by the data. Afterwards, the model was fitted for the fixed effects under consideration. The fixed effect was kept in case it shows significant results following the χ^2-test conducted for model comparisons. Results are depicted in Table 7.6. The model comparisons were conducted using a likelihood ratio test (LRT) because GLMMs can not be compared using the Kenward-Roger- or other methods (Singmann & Kellen, 2017).

Table 7.6: Hierarchical approach for crash. In case only random intercepts are supported, random slopes and the correlation between them showed convergence problems. If only random intercepts and random slopes are supported, their correlation showed convergence issues. Observations n = 299, unless specified otherwise due to missing data from the eye-tracker.

Fixed effect	Maximal random-effect structure supported by the data (no convergence problems)	Contribution of fixed effect (compared to model without it, likelihood ratio test)
	Traits	
Sex	Random intercepts: `(1|nr)`	$\chi^2(1) = .13, p = .72$
Age (linear)	Random intercepts: `(1|nr)`	$\chi^2(1) = .00, p = .98$
Mileage per year	Random intercepts: `(1|nr)`	$\chi^2(1) = .07, p = .99$
Subjective driving style	Random intercepts and slopes: `(subj. driving style||nr)`	$\chi^2(1) = .10, p = .75$
	Situational factors	
Traffic density	Random intercepts: `(1|nr)`	$\chi^2(1) = .18, p = .67$
Situation	Random intercepts: `(1|nr)`	$\chi^2(1) = 5.01, p = .08$
Situation shows only a tendency for significant effects, but is analyzed regardless. For the following models, Situation is considered as fixed effect, but is not reported again for the individual model comparisons which include models with/without "Situation" also.		
	State changes	
Type of NDRT	The GLMM shows convergence issues for all random effect structures and Type of NDRT is dropped from further analysis.	
HGD (last minute prior to RtI), (n = 290)	Random intercepts: `(1|nr)`	$\chi^2(1) = .51, p = .48$
Change of the HGD, (n = 276)	Random intercepts: `(1|nr)`	$\chi^2(1) = .19, p = .66$
PEOR (last minute), (n = 285)	The GLMM shows convergence issues for all random effect structures and PEOR(last minute) is dropped from further analysis.	
Change of the PEOR, (n = 271)	Random intercepts: `(1|nr)`	$\chi^2(1) = .00, p = .97$
PEOR (10 seconds)	Random intercepts: `(1|nr)`	$\chi^2(1) = .05, p = .83$
Blink duration, (n = 266)	Random intercepts: `(1|nr)`	$\chi^2(1) = .07, p = .80$
Change of the blink duration, (n = 252)	Random intercepts: `(1|nr)`	$\chi^2(1) = .35, p = .56$
Blink frequency, (n = 266)	Random intercepts: `(1|nr)`	$\chi^2(1) = .72, p = .40$
Change of the blink frequency, (n = 253)	Random intercepts: `(1|nr)`	$\chi^2(1) = .11, p = .74$

Change of the COP in the seat	Random intercepts: `(1|nr)`	$\chi^2(1) = .21, p = .65$
Change of the COP in the backrest	Random intercepts: `(1|nr)`	$\chi^2(1) = 1.13, p = .29$

No fixed factor shows significant results in the model comparisons as depicted in Table 7.6. A tendency is observed for the factor Situation. The results represent the crashes happened due to either random reasons or reasons not accounted for by the set of factors specified in this approach. While this conclusion seems unsatisfactory for any modeling approach, it can be explained by inspecting the total number of crashes with respect to the observations. Only nine crashes were identified in total, compared to 290 no-crash events. The number of no crashes is disproportionately outweighing the number of crashes by a ratio of approximately 32 to one. In light of this ratio, understanding the occurrence of crashes for the design of experiments in this work, any kind of GLM or GLMM is unlikely to identify the underlying reasoning for crashes.

The factor situation shows a tendency likely due to 5 out of 9 crashes happening in the crash site. Contrary, two participants accounted for a total of 6 crashes rendering any additional analysis void.

7.2.3 Modeling of time to collision

The TTC was modeled following the same way as TOT. The distribution of the data set was plotted in Figure 7.11. During the hierarchical process, two fixed effects (Traffic density and Type of NDRT) showed highly significant results when comparing the linear mixed models, but introduction led to convergence issues when fitting the model for both factors independent of the random structure. This is addressed after the final linear mixed model is introduced by an alternative linear modeling approach not incorporating random effects. Thus, the final mixed model does not include the fixed factors Traffic density and Type of NDRT.

Table 7.7: Hierarchical approach for TTC. In case only random intercepts are supported, random slopes and the correlation between them showed convergence problems. If only random intercepts and random slopes are supported, their correlation showed convergence issues. Observations n = 299, unless specified otherwise due to missing data from the eye-tracker.

Fixed effect	Maximal random-effect structure supported by the data (no convergence problems)	Contribution of fixed effect (compared to model without it, Kenward-Roger-method)
	Traits	
Sex	Random intercepts: `(1|nr)`	$F(1, 80.14) = 1.75, p = .19$
Age (quadratic)	-	$F(2, 77.97) = 3.81, p = .03$
For the following models, Age is always considered as fixed effect, but is not reported again for the individual model comparisons which include models with/without "Age(quadratic)" also.		
Mileage per year	Random intercepts: `(1|nr)`	$F(3, 75.47) = 1.35, p = .27$

Subjective driving style	Random intercepts, random slopes and their interaction: (subj. driving style|nr)	$F(1, 41.15) = 2.54, p = .12$
	Situational factors	
Traffic density	No random effects structure supported.	$F(1, 110.83) = 51.61$, $p < .001$
Fitting the model including Traffic density always yielded convergence issues and produced a singular fit when the fixed effect Traffic including random intercepts was introduced. The fixed effect yielded highly significant results and is considered in an alternative, linear model approach for future discussion. To avoid mixing/missing convergence issues for the random effects structure of other fixed effects, traffic is set back.		
Situation	Random intercepts, random slopes and their interaction: (situation|nr)	$F(2, 56.28) = 42.87$, $p < .001$
For the following models, Situation is always considered as fixed effect, but is not reported again for the individual model comparisons which include models with/without "Situation" also.		
	State changes	
Type of NDRT	No random effects structure supported.	$F(3, 243.44) = 9.50$, $p < .001$
Fitting the model including the factor Type of NDRT always yielded convergence issues and produced a singular fit when the fixed effect including random intercepts was introduced. The fixed effect yielded highly significant results and is considered in an alternative, linear model approach for future discussion. To avoid mixing/missing convergence issues for the random effects structure of other fixed effects, Type of NDRT is set back.		
HGD (last minute prior to RtI), (n = 290)	Random intercepts: (1|nr)	$F(1, 228.08) = .12, p = .73$
Change of the HGD, (n = 276)	Random intercepts: (1|nr)	$F(1, 249.39) = .62, p = .43$
PEOR (last minute), (n = 285)	Random intercepts: (1|nr)	$F(1, 236.90) = 9.51$, $p = .002$
For the following models, PEOR (last min.) is always considered as fixed effect, but will not be reported again for the individual model comparisons which include models with/without "PEOR(lastmin)" also.		
Change of the PEOR, (n = 271)	Random intercepts: (1|nr)	$F(1, 254.21) = .68, p = .41$
PEOR(10s) (n = 271)	Random intercepts: (1|nr)	$F(1, 251.52) = .35, p = .56$
Due to the high correlation between PEOR(lastmin) and PEOR(10s) (see Table B.1), an additional model was fitted only including PEOR(10s) and temporarily dropping PEOR(lastmin). The model also shows significant results for PEOR(10s).		

PEOR(10s w.o. PEOR(lastmin)) (n = 271)	Random intercepts: `(1|nr)`	$F(1, 245.35) = 6.12, p = .01$
Blink duration, (n = 266)	Random intercepts: `(1|nr)`	$F(1, 249.08) = 1.73, p = .19$
Change of the blink duration, (n = 252)	Random intercepts: `(1|nr)`	$F(1, 229.35) = .08, p = .78$
Blink frequency, (n = 266)	Random intercepts `(1|nr)`	$F(1, 157.13) = 2.30, p = .13$
Change of the blink frequency, (n = 252)	Random intercepts: `(1|nr)`	$F(1, 229.71) = 1.04, p = .31$
Change of the COP in the seat, (n = 285)	Random intercepts and random slopes: `(copseat(changes)||nr)`	$F(1, 39.48) = .008, p = .98$
Change of the COP in the backrest, (n = 285)	Random intercepts: `(1|nr)`	$F(1, 269.68) = .23, p = .64$

Table 7.7 provides the hierarchical process from top to bottom. Age modeled as quadratic term shows significant results with younger and older participants producing lower TTCs compared to - few - middle-aged participants. Only random intercepts are supported by the underlying data.

Traffic density and the factor Situation show highly significant results. The factor Situation was kept for the ongoing modeling approach while Traffic density was dropped temporarily. The Type of NDRT shows highly significant results as well but also introduced convergence issues to the model fitting and was dropped for the final mixed model.

The PEOR during the last minute of automated driving prior to a RtI shows significant results including random intercepts. A reduced model without the PEOR(lastmin) but considering the PEOR(10s) revealed significant results, underlining the correlation between PEOR(lastmin) and PEOR(10s). Combined results put emphasis on the significant effect of visual attention and the resulting criticality of a take-over displayed by the TTC. More PEOR lead to higher TTCs representing less critical take-overs.

No other factors could be identified to significantly account for variance in the outcome TTC and the final mixed model was fitted including Age, Situation and PEOR(lastmin) incorporating random intercepts as the maximal random effect structure supported by the underlying data.

Concerning the assumptions addressed at the beginning of this chapter, the residuals of the final model were plotted over the estimated values in Figure 7.12. Visual inspection does not reveal an obvious indication of heteroscedasticity, even though a slight trend towards larger spread residuals for higher values of estimated TTC is observed. Influential points are identified by examining the Cook's distance for the nested groups/participants. Results are plotted in Figure B.4. Participants with IDs 10, 21 and 24 are influential points. Regarding the DFBetas for all factors individually in Figure B.5 and the percentage in change in Figure B.6 with/without the influential participants, all three exert a large influence on the model fit. To better understand the magnitude of these influential points, an additional linear mixed model was fitted without these points and compared to the original model. The final model equations can be found in Table 7.8. The original data together with the fitted model separated for each fixed effect are plotted in Figures 7.13,

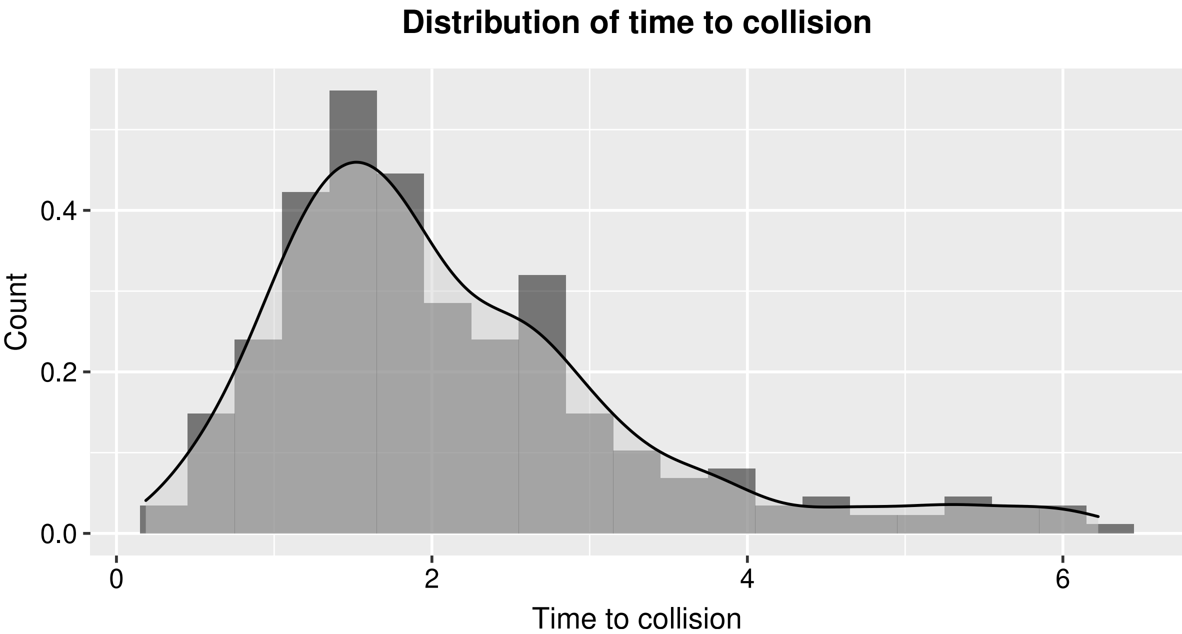


Figure 7.11: Histogram and density plot of the time to collision distribution. The linear mixed modeling approach incorporating random effects is based on visual examination of this plot.

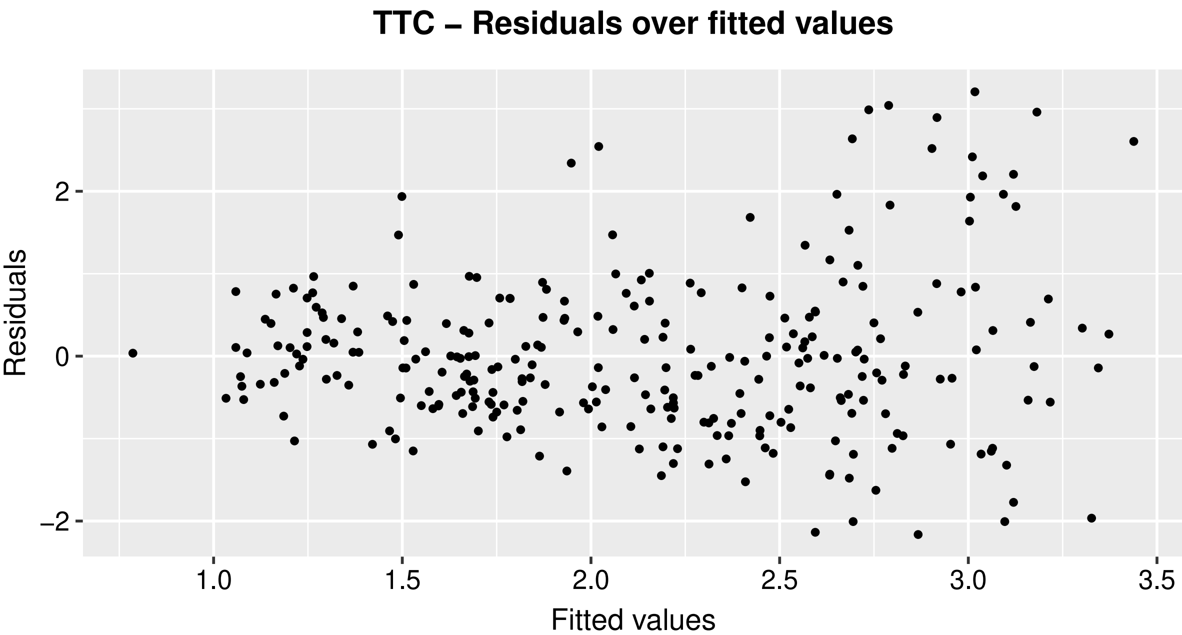


Figure 7.12: Plot of the residuals over the fitted values of the final model for the TTC. The distribution is not suggesting heteroscedasticity and the assumption of homoscedasticity is not violated.

Table 7.8: Model description of final model for TTC. The values for the fitted model without participants 10, 21 and 24 are reported in brackets. The model was fitted on 278 (269) observations.

		Model fit			
AIC	Pseudo-R^2 (fixed effects)		Pseudo-R^2 (total)		RMSE
830.56 (770.56)	R^2=.25, (.25) (fixed effects)		R^2=.29, (.26) (total)		.97 (.93)
		Fixed effects			
Factor	Estimate	Std. error	t-value	df	p-value
Intercept	1.96, (1.88)	.14, (.14)	13.51, (13.89)	204.70, (194.73)	<.001, (<.001)
Age	.32, (.29)	1.11, (1.00)	.29, (.29)	67.67, (65.44)	.77, (.77)
Age (quadratic)	-2.39, (-1.37)	1.12, (1.01)	-2.14, (-1.36)	71.91, (68.98)	.04, (.18)
Crash site	-.74, (-.70)	.09, (.09)	-8.35, (-8.14)	219.94, (217.94)	<.001, (<.001)
Interstate crossing	.20, (.22)	.09, (.09)	2.18, (2.46)	228.40, (226.44)	.03, (.01)
PEOR (lastmin)	.01, (.01)	.00, (.00)	3.11, (3.31)	230.49, (210.12)	.002, (<.001)
		Random effects			
Group	Parameter		# Groups	Std. Deviation	ICC
Participant nr.	Intercept		84 (81)	.24, (.14)	.05, (.02)
Residual	–		–	1.00, (.95)	–

Table 7.9: Correlation matrix of fixed effects for TTC. Note, that for linear mixed effect models, this matrix is "an approximate correlation of the estimator of the fixed effects" Baayen (2009).

	Intercept	Age	Age (quadr.)	Sit. crash	Sit. crossing
Age	-.05				
Age (quadr.)	-.02	.00			
Sit. crash site	-.23	.04	-.09		
Sit. int. crossing	-.26	.03	-.10	-.03	
PEOR	-.85	.04	.07	.04	.11

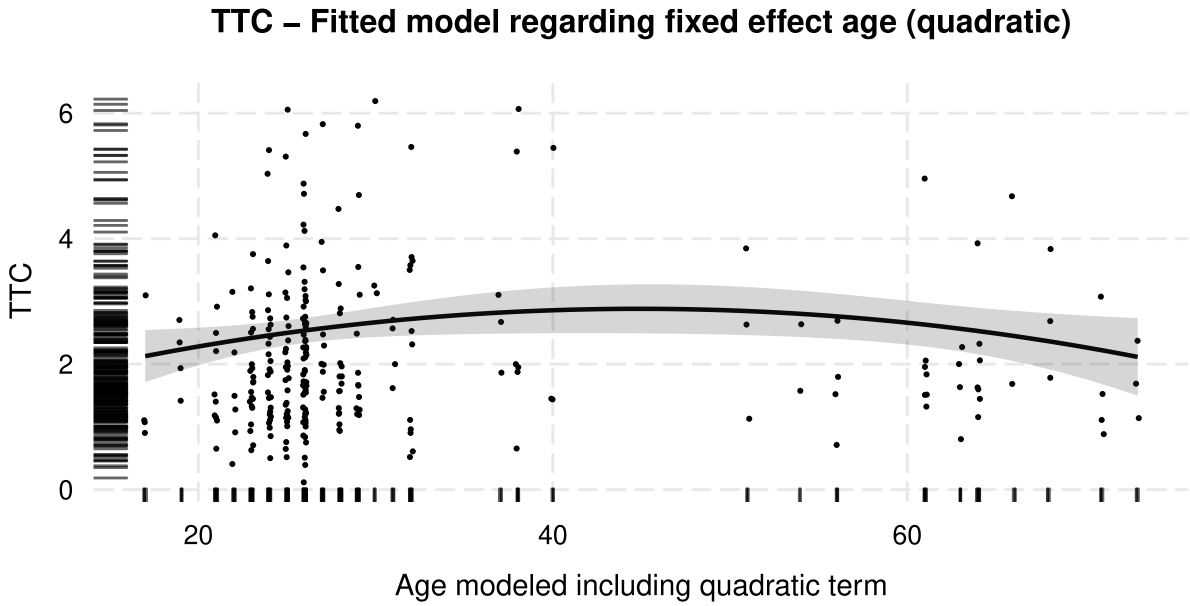


Figure 7.13: Plot of the fitted model for the fixed effect Age modeled as quadratic term. The original data are also plotted. All participants were included.

7.14 and 7.15. The final model is plotted in Figure 7.16 and shows the effect of Traffic, since only participants Nr. 1-34 could experience no traffic density, likely leading to higher TTCs.

Regarding the results for the random effects in Table 7.8, where the mixed model shows an ICC of .05 and the reduced model shows an ICC of .02, fitting a linear regression model without a random effects structure seems more feasible. The random intercepts account for some of the variance stemming from outliers or barely account for any variance at all. The approach of fitting a linear mixed model can be regarded to be too complex for the underlying data in this approach.

To allow a more comprehensive understanding of all relevant effects, Table 7.10 shows the results of an alternative linear model using ordinary least squared (OLS) regression including predictors Traffic density and Type of NDRT without random effects. The adjusted R^2 is higher compared to the best linear mixed model fit and underlines the conclusion that the introduction of random effects for TTC is cumbersome for this data set. Concluding, the individual results from the linear regression show that situational factors play the most important role in predicting TTC as measure of criticality in a take-over. Both Situation and Traffic density show highly significant contributions to the model. A tendency for significant results can be observed regarding the Type of NDRT, with no engagement in NDRTs leading to higher TTCs. The PEOR(lastmin) significantly influences the TTC, with a higher amount of visual attention towards the road leading to less critical take-overs. No additional analysis or plotting was executed for the linear regression model since this modeling approach focuses on mixed models.

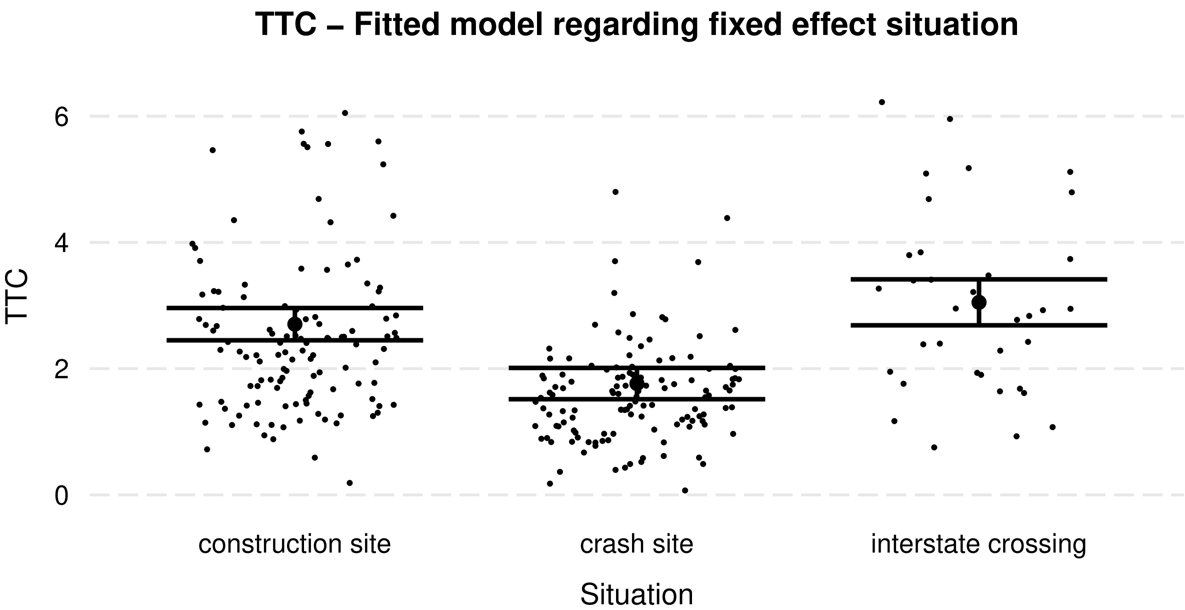


Figure 7.14: Plot of the fitted model for the fixed effect Situation only. The original data are also plotted. All participants were included.

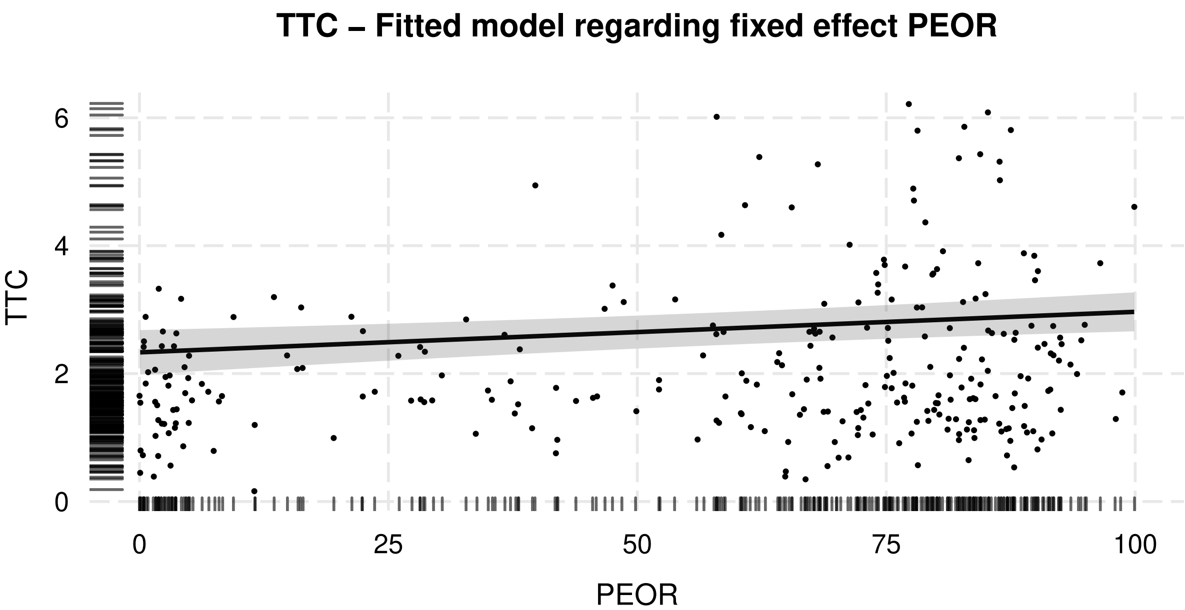


Figure 7.15: Plot of the fitted model for the fixed effect PEOR(lastmin) only. The original data are also plotted. All participants were included.

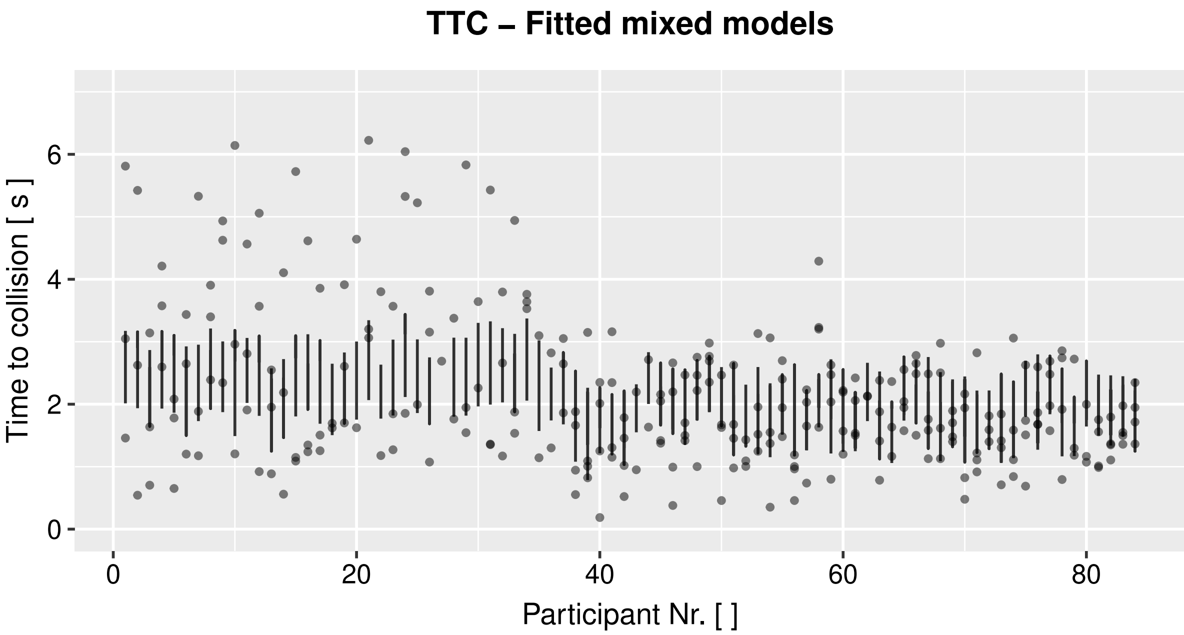


Figure 7.16: Plot of the fitted mixed models for the individual participants showing the random effects structure. The effect of Traffic (not considered in the model due to convergence issues) is apparent, since only participants Nr. 1-34 could experience no traffic density, likely leading to higher TTCs.

Table 7.10: Model description of the linear model for TTC without random effects. The model was fitted on 278 observations using OLS regression.

		Model fit		
R^2 = .36		RMSE = .94		
		Fixed effects		
Factor	Estimate	Std. error	t-value	p-value
Intercept	2.44	.26	9.22	<.001
Age (quadratic)	-.82	1.04	-.79	.43
Crash site	-.62	.09	-6.95	<.001
Interstate crossing	.35	.09	3.86	<.001
PEOR(lastmin)	.01	.00	2.15	.03
Traffic density	-.04	.01	-4.61	<.001
N-back	-.18	.15	-1.19	.23
No task	.22	.12	1.82	.07
Motoric	-.21	.12	-1.71	.09

7.2.4 Modeling of longitudinal acceleration

For the modeling of the minimal longitudinal acceleration, the same procedure as described for TOT and TTC was initially applied. The hierarchical process fitting linear mixed models and comparing them using the *mixed*-function from the *afex*-package (Singmann et al., 2019) led to convergence issues throughout for the first couple of fixed effects. Reducing the order of the random effects structure to only considering random intercepts did not solve the convergence issues. The distribution of the data was plotted in Figure 7.17 and revealed a highly non-normal distribution of data. A log-transformation for the outcome longitudinal acceleration also led to convergence problems throughout when considering only random intercepts. Examining the interval between -10 m/s^2 and -5 m/s^2 the distribution of values shows a visually identifiable normal distribution, in combination with an "almost-zero"-inflated peak of values between 0 m/s^2 and -1 m/s^2. Following a broader approach on data modeling and guidelines from Zuur and Ieno (2016), zero-inflated distributions in general can be modeled using a variety of models, such as generalized linear mixed models (GLMMs) incorporating e.g. a gamma distribution. The *afex*-package incorporates the possibility of fitting GLMMs in the same way as described for linear mixed models that were fitted for TOT and TTC, see modeling of crash probability. An approach fitting GLMMs using a gamma distribution also failed due to convergence issues when only random intercepts were considered. Regarding the plotted distribution in Figure 7.17, this can be comprehended, since the distribution shows a drop in density right after longitudinal acceleration close to zero, defying the general assumptions of a gamma distribution. Based on the results and conclusion from Gold (2016), the distribution can be accounted for by generally splitting participants' reaction behavior into braking or non-braking. The "almost-zero"-inflated distribution shows the drag moment from the vehicle after the RtI when the automation was deactivated, amounting to values of up to -.64 m/s^2. Participants who did not brake but reacted by only steering led to the skewed distribution concerning the minimal longitudinal acceleration. Participants who did brake showed a rather normal distribution around approximately -7 m/s^2. While the data set and the attempted model fit allow a detection of medium but not small effects (see Figure 7.2), splitting the data following Gold (2016) would decrease potential effect sizes.

Based on the distribution at hand and the model fitting attempts with the full data set leading to only convergence issues, a data split was evaluated on a qualitative analysis of data and the following assumptions:

1. Focus of the modeling approach is to compare the influence of different fixed effects including the idiosyncratic random effects from non-independent data.

2. Data of minimal longitudinal acceleration were split into braking or non-braking participants by distinguishing values above/below .64 m/s^2.

3. While a prior decision in "braking or no braking" can be modeled following the example of Gold (2016), the split data set in this work would be fitted for braking participants regardless of the binomial model fit.

The reduced data set for the longitudinal acceleration after the split consisted of 219 observations. The number of participants was reduced to 81. Out of the original 84 participants, three participants did not brake in any of the take-over situations but always reacted to the RtI by executing a steering maneuver. Consequently, the vast majority of participants did not always use either one maneuver to solve the take-over situation

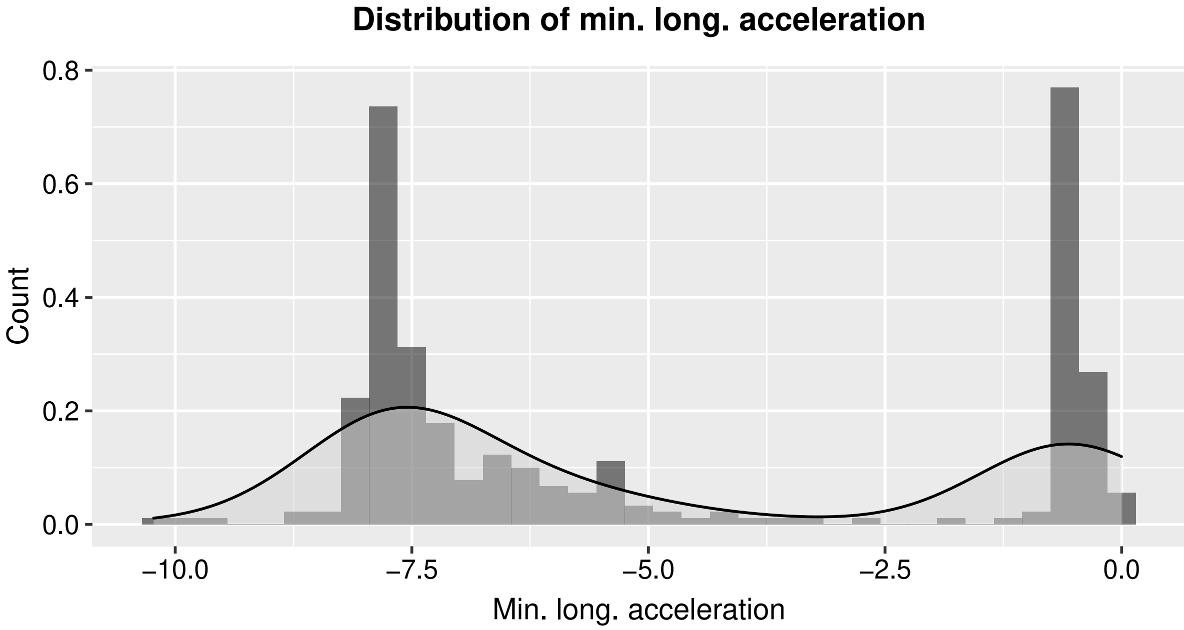


Figure 7.17: Histogram and density plot of the min. long. acceleration distribution.

but typically combined a braking and a steering maneuver. In addition with the solution from Gold (2016) where the final modeling approach consisted of a multinomial logistic regression, a potential solution for this data set consists of a multinomial mixed logistic regression (MMLR).

While there are no packages in *R* tying into the procedure of *afex* including model comparisons for MMLR, the package *ordinal* (Christensen, 2019) allows the fitting of individual MMLR models. The feasibility of adding random effects is assessed by checking each model output, specifically the reported gradient and the Hessian index as indication of successful fitting. Adaptive Laplace approximation is used to numerically provide maximum likelihood estimates for individual model parameters, i.e. fixed and random effects (Christensen, 2019). In addition, an optimization is executed by *ordinal* to obtain the model fit including the random effects. If the optimization of estimated random effects from the Laplace Approximation is successful, the gradient shows results close to zero while the Hessian index should stay below a threshold of approximately 10^5. The results after approximation also provide information on the variance-covariance matrix of parameters and report warnings in case of ill-defined matrices. Concerning the specific procedure and peculiarities of fitting MMLR models and including knowledge on the automated model comparisons as described for TOT, TTC and crash, the following procedure is implemented.

1. A basic MMLR model is determined, with no fixed effects (fitting a single global intercept) and random intercepts (since the random effects are of interest, no comparison with non-mixed multinomial logistic regression is calculated).

   ```
   a_long(categorical) ~ 1 + (1|nr)
   ```

2. A new MMLR model is fitted, including a specific fixed effect, e.g. Age. In a first step, the maximal random effects structure[4] is considered.

[4]Note: For MMLR models, the *ordinal*-package can only comply with a maximal random effects structure of random intercepts and random slopes. A potential correlation between them cannot be fitted and is not

```
a_long(categorical) ~ age + (age||nr)
```

3. In case a warning concerning an ill-defined variance-covariance matrix is issued, or the gradient does not show values close to zero, or the Hessian index is exceeding 10^5, the fitted model is determined to not be supported by the underlying data concerning the random effects structure.

4. In that case, the random effects structure is reduced to random intercepts only and the model is refitted.

```
a_long(categorical) ~ age + (1|nr)
```

5. If reported warnings or issues persist, the model for the specific fixed effect (in this case Age) does not support any random effects and the fixed effect is dropped from further analysis. In case too many fixed effects are dropped due to convergence issues, including random effects to the modeling approach does not seem feasible.

6. However, in case the model shows a converging fit, it is compared to the basic MMLR model (without any fixed effects, including random intercepts) using a likelihood ratio test (LRT). In case the provided p-value is below .05, the fixed effect is determined to significantly improve the model fit by accounting for variance in the outcome and is kept for the next step. In case the LRT reveals non-significant results, the fixed effect is dropped due to the new model not significantly accounting for variance in the outcome.

7. The procedure is executed for all fixed effects analogue to the modeling approaches of TOT, TTC and crash.

The following table depicts the results from the model comparisons with the respective fixed effects compared to the basic MMLR model.

Table 7.11: Hierarchical approach for minimal longitudinal acceleration. In case only random intercepts are supported, random slopes show convergence problems. Observations n = 299, unless specified otherwise due to missing data from the eye-tracker.

Fixed effect	Maximal random-effect structure supported by the data (no convergence problems)	Contribution of fixed effect (compared to model without it, likelihood ratio test.)
	Traits	
Sex	Random intercepts: `(1|nr)`	$\chi^2(1) = 1.24, p = .27$
Age (qudaratic)	Random intercepts: `(1|nr)`	$\chi^2(2) = 3.57, p = .17$
Mileage per year	Random intercepts: `(1|nr)`	$\chi^2(3) = 4.37, p = .22$
Subjective driving style	Random intercepts and slopes: `(subj. driving style||nr)`	$\chi^2(3) = 2.97, p = .40$
	Situational factors	

considered. Regarding the maximal random structure supported by the data following results from modeling TOT and TTC, this is evaluated as not critical.

Traffic density	Random intercepts and slopes: `(traffic||nr)`	$\chi^2(1) = .57, p = .45$
Situation	Random intercepts and slopes: `(situation||nr)`	$\chi^2(7) = 72.36, p < .001$
Situation shows highly significant effects. For the following models, Situation is always considered as fixed effect, but is not reported again for the individual model comparisons which include models with/without Situation also.		
State changes		
Type of NDRT	Random intercepts and slopes: `(ndrt||nr)`	$\chi^2(13) = 6.34, p = .93$
HGD (last minute prior to RtI), (n = 290)	Random intercepts and slopes: `(HGD(lastmin)||nr)`	$\chi^2(4) = 2.66, p = .62$
Change of the HGD, (n = 276)	Random intercepts: `(1|nr)`	$\chi^2(1) = .33, p = .56$
PEOR (last minute), (n = 285)	Random intercepts: `(1|nr)`	$\chi^2(1) = .00, p = .95$
Change of the PEOR, (n = 271)	Random intercepts: `(1|nr)`	$\chi^2(1) = .19, p = .66$
PEOR (10 seconds)	Random intercepts: `(1|nr)`	$\chi^2(1) = 2.00, p = .16$
Blink duration, (n = 266)	Random intercepts: `(1|nr)`	$\chi^2(1) = .28, p = .60$
Change of the blink duration, (n = 252)	Random intercepts: `(1|nr)`	$\chi^2(1) = .01, p = .94$
Blink frequency, (n = 266)	Random intercepts and random slopes: `(blink frequency(lastmin)||nr)`	$\chi^2(4) = .02, p = .99$
Change of the blink frequency, (n = 253)	Random intercepts: `(1|nr)`	$\chi^2(1) = .08, p = .77$
Change of the COP in the seat	Random intercepts: `(1|nr)`	$\chi^2(1) = .09, p = .77$
Change of the COP in the backrest	Random intercepts: `(1|nr)`	$\chi^2(4) = .31, p = .99$

The results in Table 7.11 show highly significant results for the situation, including random intercepts and random slopes. For the crash site, participants braked stronger in comparison to the construction site and the interstate crossing whilst showing strong individual differences both concerning their idiosyncratic braking behavior (random intercepts) and their change in behavior depending on the situation at hand (random slopes). The estimated probabilities of the longitudinal accelerations are plotted in Figure 7.18 and underline the results. The final model is described in Table 7.12. Comparing both the Pseudo-R^2 for fixed effects only and the full model and the ICC, the random effects explain half of the variance in the outcome. Participants tend to either brake or not brake, depending on their individual disposition. This is moderated to a small degree by the situation, while the high variance of random effects for the crash site can be understood as to motivate some participants to brake, while others do not brake in the crash site. Overall,

Table 7.12: Model description of the final model for categorical longitudinal acceleration. A multinomial logistic mixed model was fitted using Laplace approximation on 299 observations.

		Model fit		
AIC	Pseudo-R^2 (fixed effects)		Pseudo-R^2 (total)	
559.75	R^2 = .17 (fixed effects)		R^2 = .67 (total)	
		Fixed effects		
Factor	Estimate	Std. error	z-value	p-value
Crash site	-2.59	.61	-4.28	<.001
Interstate crossing	-.02	.43	-.04	.97
		Random effects		
		ICC = .60		
Group	Parameter	# Groups	Variance	Std. Deviation
Participant nr.	Intercept	84	1.13	1.06
Participant nr.	Crash Site	84	6.43	2.54
Participant nr.	Interstate Crossing	84	.58	.76

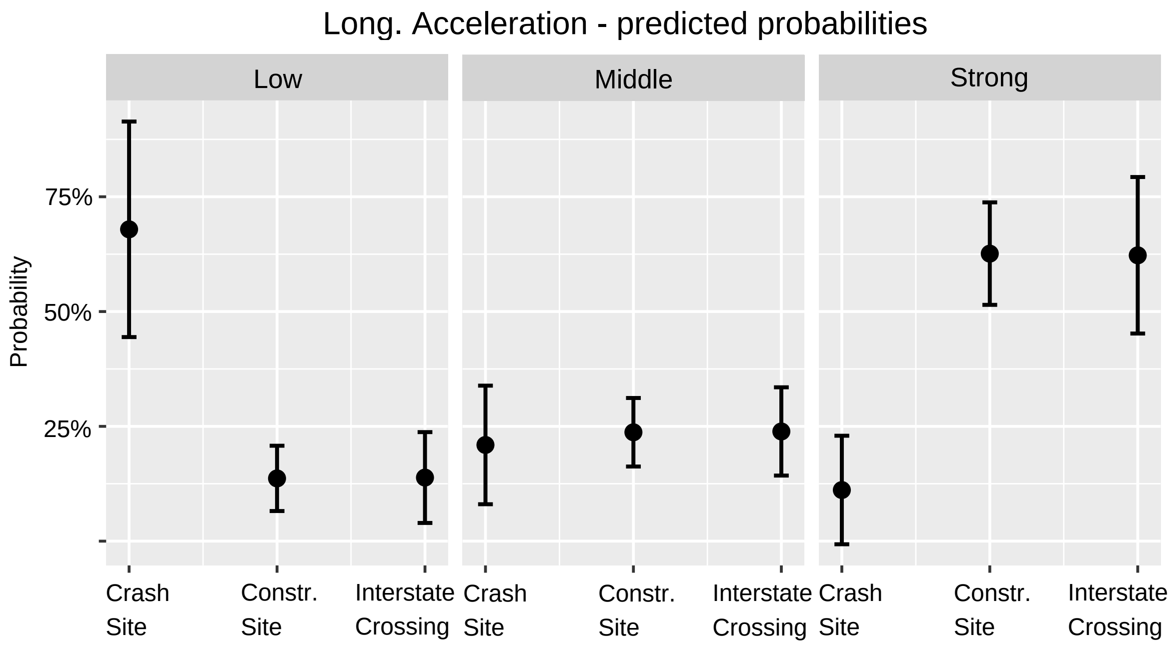


Figure 7.18: Plot of the fitted model split for the effect situation including 95% confidence interval.

the crash site promotes low longitudinal acceleration whereas participants tend to brake stronger in the construction site and the interstate crossing. Regardless, the individual tendency to brake in a take-over or not is highly dependent on the specific participant at hand. The final model accounts for two thirds of the total variance in the outcome, providing a much better prediction quality compared to TOT and TTC. No state changes showed significant contributions to the final model, adding to the exceptional influence of individual dispositions concerning braking behavior.

7.2.5 Modeling of lateral acceleration

The modeling of the lateral acceleration initially followed the depicted process for TOT and TTC with regard to (Gold, 2016). The histogram and density plot depicted in Figure 7.19 showed a highly skewed "almost-zero"-inflated distribution. From visual examination, the maximal lateral accelerations are not normally distributed. This can be attributed to the take-over situations applied in the experiments. In Experiment 1, one out of three situations consisted of the construction site. In Experiment 2, two out of four take-overs were in the construction site. Thus, a substantial number of take-overs were experienced in a situation that did not require a lane change after taking over. Lateral accelerations of close to zero could represent the number of participants exerting only small accelerations when stabilizing the vehicle. Following the hierarchical process, analogue to the initial process concerning the longitudinal accelerations, the introduction of only random intercepts showed consistent convergence issues fitting linear mixed models. To avoid these convergence issues, the process was also adapted to incorporate the fitting of GLMMs with a gamma distribution, showing more similarity to the distribution in Figure 7.19. The specification of random intercepts also led to convergence issues throughout. A log-transformation of the outcome in addition with incorporating a gamma-distributed GLMM modeling approach also showed convergence issues. The steps described were chosen to both generate fitting models including a random effects structure and allow an easy-to-interpret solution.

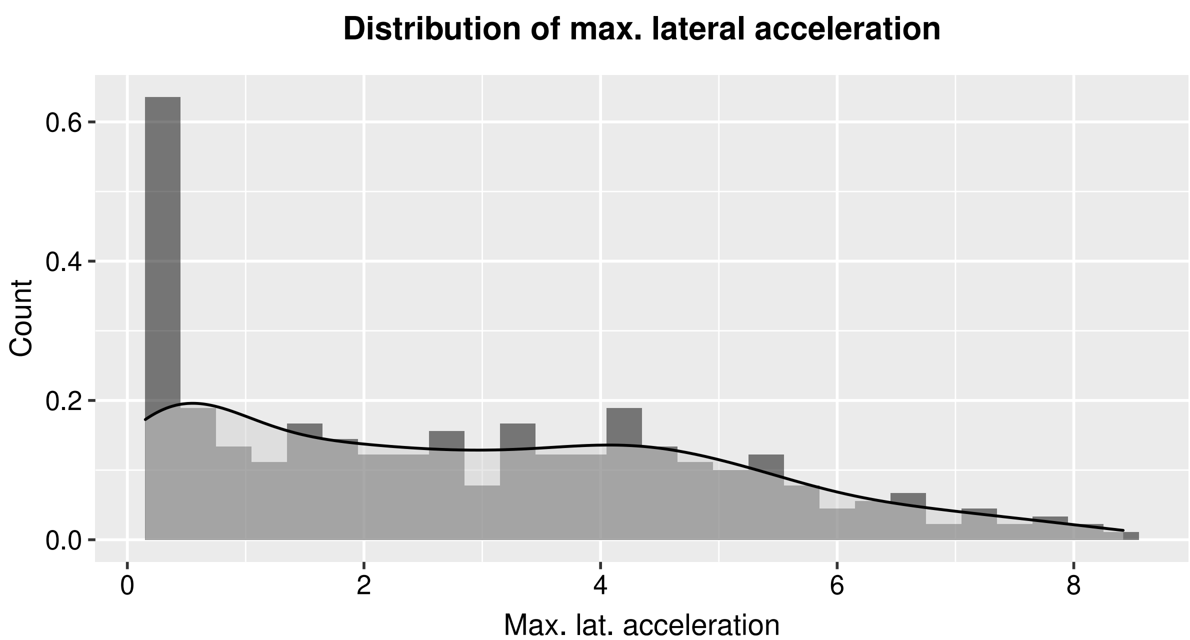


Figure 7.19: Histogram and density plot of the max. lat. acceleration distribution.

While the distribution of the longitudinal accelerations called for a MMLR approach more clearly, the persisting convergence issues led to a split of lateral acceleration into three groups, low, medium and strong accelerations. MMLR models were fitted for the lateral accelerations utilizing the same approach as depicted for the longitudinal ones.

Table 7.13: Hierarchical approach for maximal lateral acceleration. In case only random intercepts are supported, random slopes show convergence problems. Observations n = 299, unless specified otherwise due to missing data from the eye-tracker.

Fixed effect	Maximal random-effect structure supported by the data (no convergence problems)	Contribution of fixed effect (compared to model without it, likelihood ratio test)
	Traits	
Sex	Random intercepts and slopes: `(sex||nr)`	$\chi^2(4) = 2.95, p = .57$
Age (qudaratic)	Random intercepts and slopes: `(age||nr)`	$\chi^2(8) = 6.62, p = .58$
Mileage per year	Random intercepts and slopes: `(km(peryear)||nr)`	$\chi^2(13) = 3.18, p = .99$
Subjective driving style	Random intercepts and slopes: `(subj. driving style||nr)`	$\chi^2(4) = .81, p = .94$
	Situational factors	
Traffic density	Random intercepts and slopes: `(traffic||nr)`	$\chi^2(4) = 15.06, p < .01$
Traffic shows significant effects. For the following models, Traffic is always considered as fixed effect, but is not reported again for the individual model comparisons which include models with/without Traffic also.		
Situation	Random intercepts: `(1||nr)`	$\chi^2(2) = 165.15, p < .001$
Situation shows highly significant effects. For the following models, Situation is always considered as fixed effect, but is not reported again for the individual model comparisons which include models with/without Situation also.		
	State changes	
Type of NDRT	Random intercepts: `(1|nr)`	$\chi^2(3) = 3.42, p = .33$
HGD (last minute prior to RtI), (n = 290)	Random intercepts: `(1|nr)`	$\chi^2(1) = .22, p = .64$
Change of the HGD	–	–
Fitting the model including the change of the HGD always yielded convergence issues and produced a singular fit when the fixed effect HGD(change) including random intercepts was introduced. The fixed effect yielded no results for estimated coefficients and was dropped.		
PEOR (last minute), (n = 285)	Random intercepts: `(1|nr)`	$\chi^2(1) = 2.21, p = .14$

Change of the PEOR	–	–
Fitting the model including the change of the PEOR always yielded convergence issues and produced Hessian indices exceeding the threshold when the fixed effect PEOR(change) including random intercepts was introduced. The fixed effect yielded no results for estimated coefficients and was dropped.		
PEOR (10 seconds)	–	–
Fitting the model including the PEOR in the last ten seconds before the RtI always yielded convergence issues and produced Hessian indices exceeding the threshold when the fixed effect PEOR(10s) including random intercepts was introduced. The fixed effect yielded no results for estimated coefficients and was dropped.		
Blink duration, (n = 266)	Random intercepts and slopes: `(blinkdur(lastmin)||nr)`	$\chi^2(4) = 1.49, p = .83$
Change of the blink duration, (n = 252)	Random intercepts and slopes: `(blinkdur(change)||nr)`	$\chi^2(4) = 1.06, p = .90$
Blink frequency, (n = 266)	Random intercepts and random slopes: `(blink frequency(lastmin)||nr)`	$\chi^2(4) = 1.34, p = .85$
Change of the blink frequency, (n = 253)	Random intercepts: `(1|nr)`	$\chi^2(1) = .34, p = .56$
Change of the COP in the seat	–	–
Fitting the model including the change of the COP in the seat always yielded convergence issues and produced Hessian indices exceeding the threshold when the fixed effect COP(seat) including random intercepts was introduced. The fixed effect yielded no results for estimated coefficients and was dropped.		
Change of the COP in the backrest	Random intercepts: `(1|nr)`	$\chi^2(1) = 3.16, p = .08$

Results can be found in Table 7.14 and show highly significant results for the predictors situation and traffic density. Strong lateral accelerations are unlikely in all situations, while the crash site promotes higher lateral accelerations compared to the construction site and the interstate crossing. Traffic in comparison to no traffic motivates higher lateral accelerations but to a much smaller extent compared to different situations. Contrary to the longitudinal accelerations, the lateral accelerations showed a much smaller Pseudo-R^2 for the random effects, underlined by a small ICC. Random intercepts and slopes were supported, but lateral accelerations show little dependence on idiosyncratic effects of different participants. During the modeling process, four fixed effects, after traffic and situation was accounted for, including random effects, showed convergence problems throughout and were dropped. Analogue to the modeling process of TTC, the results hint to a modeling approach without random effects. Figure 7.20 shows the effect of the fixed effects traffic and situation.

Table 7.14: Model description of the final model for categorical lateral acceleration. A multinomial mixed logistic model was fitted using Laplace approximation on 299 observations.

		Model fit		
AIC	Pseudo-R^2 (fixed effects)		Pseudo-R^2 (total)	
288.03	R^2 = .56 (fixed effects)		R^2 = .61 (total)	
		Fixed effects		
Factor	Estimate	Std. error	z-value	p-value
Traffic	.09	.03	3.21	.001
Crash site	4.23	.57	7.45	<.001
Interstate crossing	1.25	.64	1.95	.05
		Random effects		
		ICC = .09		
Group	Parameter	# Groups	Variance	Std. Deviation
Participant nr.	Intercept	84	.61	.78
Participant nr.	Traffic	84	.01	.07

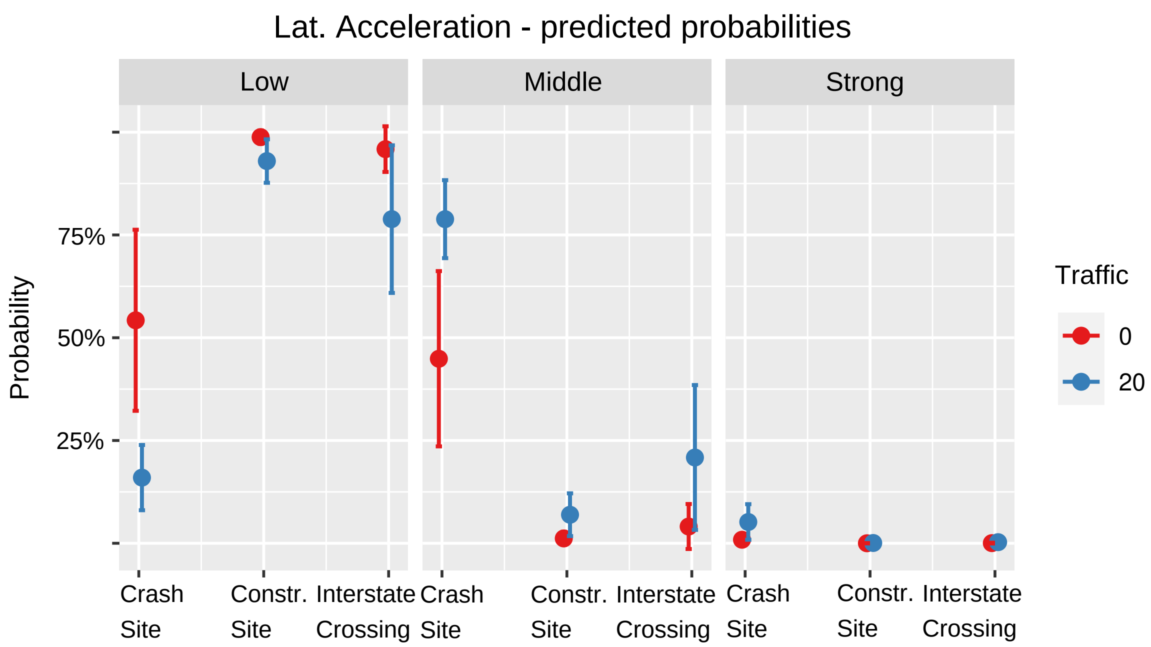


Figure 7.20: Plot of the fitted model for categorical lateral acceleration split for the effects traffic and situation including 95% confidence interval.

7.3 Discussion

The results from the modeling approach provide valuable insight into the influence of specific fixed and random effects on both time and quality aspects of take-over performance combining the data from Experiments 1 and 2.

The introduction of random effects allows the quantification of idiosyncratic effects from individual participants and shows large differences between metrics. The prediction of categorical brake accelerations and the TOT benefited whereas the TTC and categorical lateral accelerations hint towards modeling approaches without random effects.

Interpreting the random effects, the TOT seems to be influenced by the individual quickness of taking over. Participants tend to take-over faster or slower while the individual quickness is not influenced differently by fixed effects (no random slopes). The effect of these random intercepts remains small compared to the remaining residuals and fixed effects such as different situations, which show the most dominant effect on TOT. A potential explanation could lie in regarding participants to either act more spontaneous or not, or participants comprehending the situations faster than others. Regardless of underlying reasons, the individual differences between participants are important to consider in safety relevant questions, since the maximal human performance concerning TOT is dependent on both fixed and random effects. While the final model for TOT falls short of improving prediction quality compared to Gold (2016), it quantifies the importance of including random effects for TOT.

The braking behavior is highly affected by the individual disposition of participants, accounting for half of the total variance in the outcome. In addition, random slopes showed a significant model contribution. This translates into a deeply routed tendency to brake strong or brake little, moderated differently depending on the specific take-over situation for individual participants. Braking decreases the dynamic of the vehicle and thus of the take-over situation, an effect utilized in automated brake applications in take-overs (Gold, Lorenz, & Bengler, 2014). Linking the results for TTC and brake acceleration, the highly significant contribution of random effects does not seem to lead to less critical TTCs. The TTC model does not benefit from the introduction of random effects. In addition, participants that do not brake due to their internal tendency to do so, are not necessarily compensating by stronger steering maneuvers. While low braking is highly probable in the crash site and strong braking can bee seen in the construction site and the interstate crossing due to individual dispositions, the effect of lateral accelerations from steering is strongly linked to the situation and the traffic density but not to random effects. The reasoning behind these differential findings is likely found in the take-over situation design. The need for steering is clearly comprehensible based on surrounding traffic and the situation, whereas braking was rather motivated by individual risk classification or disposition.

Future modeling attempts should include a larger number of take-overs per participant to better estimate the random effects structure. The significant contribution of random effects in this chapter is based on either 3 or 4 take-overs per participants, presumably revealing medium to strong effects. Random effects should ideally be estimated based on at least 10-20 observations (Meteyard & Davies, 2019) to allow a better estimation, potentially revealing additional medium or small individual differences. While both random intercepts for TOT and random slopes for the brake accelerations were found to represent the underlying random effects structure for this data set, this should not be considered to actually represent the maximal supported random structure.

The chosen hierarchical process, consisting of a step-wise forward process, is contrary to most common practice advice from Singmann and Kellen (2017) or Meteyard and Davies (2019). The process was chosen to pinpoint convergence issues stemming from the introduction of more complex random structures such as random slopes associated with specific fixed effects. The drawback of this approach is potentially missing fixed effects added too early to the model comparisons. Future research on predicting take-over performance should include the findings on the maximal supported random effects structure from this work, but start with the maximal number of fixed effects rather than considering a step-wise forward method.

The TTC as measure of criticality shows no improvement from introducing random effects. Concluding the random effects, an assessment of take-over performance based on TTC alone would seem to show dominant situational effects such as traffic density or different situations, regardless of individual differences in take-over behavior concerning reaction time and decision on maneuver type (lateral acceleration). While a comprehensive understanding of take-over performance should include random effects, overall findings concerning situational effects are in line with literature findings (Gold, 2016; Lu et al., 2017; Zhang et al., 2019) indicating that the specific take-over situation is of highest interest to determine human behavior in a take-over.

Concerning all other fixed effects, only age and PEOR show significant contributions in the TOT and TTC models. The effect of age was expected (Gold, 2016), but remains small in comparison to situational effects, also in line with findings from literature (Körber, Gold, Lechner, & Bengler, 2016b). The group of participants consisted of mainly younger drivers, reducing a meaningful interpretation of age. In addition, while older participants show smaller TOTs, they seem to take longer to choose and execute a reaction following smaller TTCs.

More importantly, out of all "driver state fixed effects", only a high PEOR decreases TOT and increases TTC, leading to less critical reactions. The findings highlight the paradox of humans acting as fall-back in CAD: visual attention towards the road is beneficial concerning a take-over, while CAD incorporates the engagement in - often visual - NDRTs. Cognitive NDRTs can have similar effects on take-over performance in critical situations (Radlmayr et al., 2014), but visually demanding NDRTs negatively effect take-over performance in well-known situations (Gold et al., 2015) . Since the PEOR shows a significant influence in combination with the possibility of participants "looking but failed to see" (Helton & Warm, 2008), a potential consideration of mixed effects for TTC based on a larger number of take-overs per participants could reveal additional results: a tendency for significant TTC model contributions linking the type of NDRT with the TTC as criticality measure of take-overs could pave the way to a broader understanding.

Overall, the amount of variance accounted for by the final models in this work fell short from improving prediction quality compared to Gold (2016). While the combination of fewer data points and a large number of evaluated predictors can be attributed to this, the final models from this approach are deemed not feasible for potential use in future product applications. Prediction of crash probability, as most severe measure of take-over quality, did not lead to a feasible model fit due to nine crashes being disproportionately outweighed by 1:32 in relation to all recorded take-overs. These results affirm the importance of addressing specific research questions in the spectrum between maximal human performance in highly critical situations and addressing human factors questions in more frequent, less-critical take-overs.

7.4 Summary and conclusion

This chapter offers a first look on quantifiable idiosyncratic effects from individual participants in predicting take-overs but the evaluated random effects are not exhaustive considering both outcomes and predictors in take-over research. Closing with a look on the research question of the modeling approach in Figure 7.1, it can be concluded that situational factors must be addressed in any approach to shape and improve human reactions in a take-over. Individual differences of drivers should be accounted for concerning the prediction of TOT and brake accelerations, whereas TTCs and lateral accelerations seem unaffected of individual differences, at least for this data set.

The findings concerning visual attention being beneficial for both time and quality aspects of the take-over performance in addition with the dominant effect of the take-over situation motivates the focus of Experiments 3 and 4. Experiment 3 focuses on the possibility of providing peripheral monitoring of the traffic scenery while participants where engaged in a visual NDRT. Experiment 4 underlines the importance of fostering the take-over by providing additional information on the specific take-over situation at hand. While Experiment 3 in Chapter 8 focuses more on the safety relevant aspect of visual attention during CAD, Experiment 4 in Chapter 9 addresses comfort-related aspects during the take-over process.

8 Peripheral monitoring of traffic in conditionally automated driving

The experiment[1] assessed the potential benefits of peripheral monitoring of traffic on take-over performance by moving the location of a visual NDRT to the head-up display (HUD). The experiment was pre-published to this thesis at the Human Factors and Ergonomics Society (HFES) 62nd International Annual Meeting in Philadelphia (Radlmayr, Brüch, et al., 2018). This chapter provides a comprehensive overview of the experiment including main results and conclusions and is taken in part from the publication. For additional details refer to Radlmayr, Brüch, et al. (2018). Based on the results from Experiments 1 and 2 and the modeling approach in Chapter 8, this experiment focuses on the general conflict of aims in CAD. Drivers can engage in visually demanding NDRTs even though driver performance (TOT and TTC) in take-overs benefits from visual attention still focused on the surrounding traffic providing sufficient situation awareness (SA).

> "We hypothesized that drivers engaging in visually demanding NDRTs have the potential – to a certain extent – to still perceive information peripherally. The previous studies on take-over situations offered visual NDRTs displayed in the center console, making it highly unlikely to perceive surrounding traffic peripherally or at all. By moving the NDRT to the head-up display, we could simulate the engagement of drivers in a visual task in addition to allowing peripheral monitoring of traffic" (Radlmayr, Brüch, et al., 2018).

Analysis of SA during automated driving and of the take-over performance should reveal if take-overs benefit from peripheral monitoring. A semi-transparent balloon game in the HUD operationalized the visually engaging NDRT including the possibility of peripheral monitoring. The SuRT was assessed to be less feasible for operationalization as visual task in the HUD since the overall number of objects in the display of the SuRT (see Figure 6.2) would presumably clutter the HUD. An expert usability study of the balloon game prior to the main experiment underlined the feasibility of the game. Participants who could not monitor the surroundings were driving through simulated heavy fog that only allowed sight

HMI

Experiment 3: Effect of peripheral monitoring during CAD

Pre-published in Radlmayr, Brüch, et al. (2018).

RQ6 How can the HMI for CAD and the take-over be optimized by enabling the possibility of peripheral monitoring while engaging in NDRTs during CAD?

a) Does the situation awareness of drivers differ during CAD depending on the possibility of peripheral monitoring?

b) Does the potential effect of situation awareness differences affect the take-over performance?

Chapter 8

Figure 8.1: Main research question of Experiment 3.

[1] The experiment was designed and conducted with the assistance of Karin Brüch, Kathrin Schmidt, Christine Solbeck and Tristan Wehner as part of their student project (Brüch, Schmidt, Solbeck, & Wehner, 2017).

up to 10-15 meters. A third group was set to act as baseline without any NDRT and no fog, representing the possibility to monitor surroundings freely. The resulting between-subject design with one factor and three levels (no monitoring, peripheral monitoring, free monitoring) included 57 participants. The take-over situation was a construction site, known from Chapter 4, including a lane change and no additional traffic. Following the taxonomy from Gold et al. (2017), the overall criticality of the take-over situation can be classified to be low to medium. During automated driving, eye-tracking and the situation awareness global assessment technique (SAGAT) (Endsley & Garland, 2000; Franz et al., 2015) were used to measure SA. Take-over performance was analyzed using time and quality aspects identical to measures in Experiments 1 and 2. SA during the take-over process was measured using the situation awareness rating technique (SART) and eye-tracking. Measures of SA are chosen to differentiate between visual attention likely implicating perception of the surroundings, and a potential "looking but failed to see" effect (Helton & Warm, 2008) when monitoring peripherally. While PEOR could not be utilized due to overlapping areas of interest in the HUD for the NDRT and the road scenery, the HGD is used to allow an assessment of tracking behavior.

> "Results show that the free monitoring group, representing the no NDRT-condition had the highest situation awareness in the SAGAT. This is backed up by the results from the [horizontal gaze dispersion] (HGD). Participants who were not engaged in a visually demanding NDRT would use their attentional resources to scan the surroundings more often and more effectively. Consequently, they have a higher situation awareness. Regarding the take-over performance metrics, this additional situation awareness does not translate into better or faster reactions by drivers. Results from the analysis of blink frequency during the automated drive show that the game was visually demanding and was feasible as NDRT with the possibility of monitoring peripherally. Findings show that eye-tracking metrics like [the] HGD can be used to indirectly measure drivers' SA if direct measures such as SAGAT or SART are not applicable" (Radlmayr, Brüch, et al., 2018).

Significantly lower longitudinal accelerations for the no monitoring group were observed. This can be justified by the design of the take-over scenario:

> "The heavy fog cleared very quickly just prior to the request to intervene (RtI). Right after, the RtI was presented and participants had to regain manual control and execute a lane change maneuver. The combination of the clearing fog and the RtI led to a shock reaction. Comparing the longitudinal acceleration with the take-over times, the no monitoring group did not react faster but initially braked harder compared to the other groups. Participants in free and peripheral monitoring groups solved the situation by building up enough situation awareness after the RtI to execute the lane change with little braking. It can be concluded that the very low longitudinal accelerations for no monitoring are mainly due to the initial shock reaction and do not represent differences due to prior monitoring effects" (Radlmayr, Brüch, et al., 2018).

This reasoning is underlined by qualitative analysis of the raw eye-tracking videos including the head-box. It can be concluded that visual NDRTs during automated driving degrade drivers' situation awareness as expected. Peripheral monitoring by displaying the NDRT in the HUD cannot compensate this loss of SA. The differences in SA during automated driving did not lead to differences or a degradation of take-over performance.

> "Drivers were still able to build up the necessary situation awareness for a successful take-over. This implicates, that peripheral monitoring in conditionally automated driving is not necessarily recommendable because it does not result in better take-overs. Even with visual NDRTs in the head-up display, take-over performance was not degraded much compared to free monitoring" (Radlmayr, Brüch, et al., 2018).

It can be argued that the drivers in the free monitoring group did not actually use their attentional resources to monitor since their mean blink frequency was approximately the same compared to people not engaged in any (visually engaging) task (Bentivoglio et al., 1997). In addition, the methods used to measure situation awareness, SAGAT and SART, cannot be understood to measure the "same SA" (Endsley, 1998; van den Beukel & van der Voort, 2014). Further insight into the effect of different levels of situation awareness on take-over performance could benefit from using advanced methods to measure SA (Sirkin, Martelaro, Johns, & Ju, 2017) and should be pursued in future research.

Regarding the research questions in Figure 8.1, an optimization of the HMI by providing a different location of display for visual NDRTs was not successful. Experiment 4 focuses on supporting drivers during the take-over itself by providing additional information on the most dominant effect on take-over performance, the situation at hand.

9 Design and evaluation of an optimized human-machine interface for the take-over

The fourth experiment[1] aims at optimizing the HMI during the take-over process based on results from Experiments 1, 2 and the modeling approach.

So far, the experiments in this thesis focused on the maximal human performance of drivers as fallback level in CAD. Considering the subjective ratings of the take-overs in chapters 5 and 6, it becomes apparent that drivers can experience take-overs as overall inconvenient. Perceived criticality and complexity of the situations in combination with feeling time pressure result in an overall low subjective rating score for some of the take-overs. The focus on the limit of human capability to act as fallback in CAD results in highly critical and complex take-over situations. Following the taxonomy of take-over situations from Gold et al. (2017), different research emphasis require different situations. In addition to the situations , the HMI conveying the RtI in Chapters 5, 6, and 8 consists of the generic HMI requirement depicted in Chapter 4.

In order to make CAD appealing and comfortable for drivers, the role of the HMI negotiating the take-over process must be considered. Chapter 2.2.4 provides an overview of known effects on take-over performance and subjective ratings from drivers in the context of CAD. While the provided literature reveals the importance of considering the HMI for take-overs in CAD, the HUD in combination with contact-analogue information is rarely used. Lorenz et al. (2014) compared a red to a green carpet during the take-over to convey semantic information on possible corridors for drivers. The concept of the red carpet had some drivers braking to a full stop in comparison to no one braking to a full stop with the green carpet. No differences concerning the reaction times were found. In addition, analysis of the trajectory revealed better results for proposed corridors indicated by the green carpet (Lorenz et al., 2014).

Eriksson et al. (2019) used a carpet to help with the decision making process during take-overs. An improvement in correct decisions (brake or lane change maneuver) was found for the carpet or arrow interface (Eriksson et al., 2019). Regarding the differences in predicting longitudinal and lateral accelerations concerning individual contributions (random effects) in Chapters 7.2.4 and 7.2.5, additional visual information is hypothesized to help drivers in take-over situations. Future cooperative systems were identified to require multimodal and [situation] adaptive HMIs (Walch et al., 2017). Following the significant differences in take-over performance due to different take-over situations, an augmented and adaptive HMI placed in the HUD should be beneficial for the subjective rating of experienced take-overs in general. This is supported by results from Prasch and Tretter (2016) who found that a posteriori explanations for a take-over are beneficial for the psychological needs of drivers. Most importantly, a potentially degraded feeling of security caused by a take-over without obvious reasons can be mended by providing the reason for the take-over (Körber, Prasch, & Bengler, 2018).

Based on the aforementioned reasoning, this experiment is designed to develop and evaluate a HMI that supports drivers during a take-over process by providing the reason for the take-over and additional information on the exact location of the system limit and the

[1] The experiment was designed and conducted with the assistance of Sarah Werner as part of her Master's thesis (Werner, 2018)

HMI

Experiment 4: Optimization of the HMI for the take-over

RQ7 How can the HMI be optimized by offering additional information on the specific situation during the take-over process?

a) Does offering additional information in the HUD during the take-over affect the take-over performance?

b) How does the effect from additional information compare to the effect of different situations?

Chapter 9

Figure 9.1: Research questions of Experiment 4 focusing on optimized the HMI during the take-over process.

remaining distance towards it. Focus is put on the subjective assessment of the take-overs with and without the HMI and the subjective rating of the HMI concerning usability, intention of use, perceived safety and purchase intention.

9.1 Research questions

The hypothesis' for the experiment are derived from the main research questions and are not listed explicitly here. Figure 9.1 provides an overview on the underlying research questions derived from the literature review and results from previous experiments and the modeling approach.

9.2 Method

Sample

43 participants took part in the experiment. Due to technical issues with the eye-tracking system, three had to be excluded from analysis. The remaining 40 participants (25 males and 15 females) were between 21 and 75 years old, with a mean age of 30.7 years (SD = 13.0) and a median of 26 years. All participants held a regular driver's license with a mean time of possession of 13.3 years (SD = 12.1). Twenty-two participants reported to have experienced CAD in a simulator setup at least once. Fifteen participants reported to have impaired vision which was either corrected by glasses or contact lenses.

Experimental Setup

The experiment was also conducted in the static driving simulator of the Chair of Ergonomics of the Technical University of Munich. The design of experiment consisted of a mixed setup which featured two independent variables. Table 9.1 shows the two factors, their levels and the integration into the experimental setup. The within subjects factor Obviousness of the RtI was split into two levels, low and high obviousness. High obviousness was represented by the construction site, which was already used in Chapters 5, 6 and 8. It was hypothesized that a moving construction site is not likely to be integrated into a high precision map of the track and thus represents a sudden new situation for the automation that is only known from the onboard sensors. While the caution sign vehicle to the right might be classified correctly by advanced onboard systems, the exact position of potential

Table 9.1: Overview of the two independent variables and the resulting mixed design of experiment.

Factor	Levels	Design
Visual support in the take-over situation	Two (No display in the HUD, HMI in the HUD)	Between subjects factor
Obviousness of the RtI	Two (situation with high obviousness (construction site) and with low obviousness (tight curve in heavy rain))	Within subjects factor

hazards is unknown to the system due to the moving character of the situation. A plausible consequence would be the issue of a RtI. For drivers experiencing the situation, the clearly visible caution sign vehicle represents the main reason for a fast understanding of the reason for the RtI. Thus, the obviousness of the situation and the RtI is high for drivers.

A narrow curve in heavy rain represents the low obviousness condition. In dry conditions, a conditionally automated vehicle can be expected to handle the situation easily. Following the results from the project *PEGASUS* (Mazzega & Schöner, 2017), sudden changes in the friction between road and vehicle and bad weather conditions are system limits for CAD that would result in issuing a RtI. The combination of a – for interstate conditions – rather narrow curve together with the onset of heavy rain is the reason for the RtI in the situation depicted in Figure 9.3. While the automation can rely on its onboard sensors to accurately depict the friction between road and vehicle, rain and curves in general do not represent system limits. Drivers experiencing such a system could not easily predict the reason for the RtI since they have no precise understanding of the necessary requirements of the automation. A RtI in this situation represents a low obviousness of the RtI. Concerning the taxonomy of take-over situations (Gold et al., 2017), both situations generally consist of low criticality.

The factor visual support in the take-over situation was split into two levels concerning the visual display of information in the HUD. One group did not receive any visual display in the HUD. The other group received additional information in the HUD by the display of the road sign representing the reason for the RtI and the remaining distance to the system limit which was represented by two contact-analogue arrows, on either side of the road. Figures 9.2 and 9.3 show the two situations for the group with the additional HMI. The group without the additional information did not have anything displayed in the HUD. The RtI consisted of the acoustic double beep as described in Chapters 4 and the generic HMI in both groups. The development of the optimized HMI in the HUD was result of an iterative process. Concerning the modality, augmented, visual information provides the best possibility of presenting additional semantics concerning speed of perception and density of information without cluttering. The underlying conflict of limited resources of drivers that are already occupied with regaining SA and taking over needs to be considered as well: Additional information should not lead to a deterioration of the take-over performance. The literature provided at the beginning of this chapter and in Chapter 2.2.4 emphasized the feasibility of using augmented visual information.

Concerning the location of the additional information, the HUD was chosen due to the major benefit of being in line with the primary line of sight of participants taking over. Gold

Figure 9.2: The construction site with the additional HMI at the moment of the RtI (left) and 33 meters away from the system limit (right) viewed through the windshield.

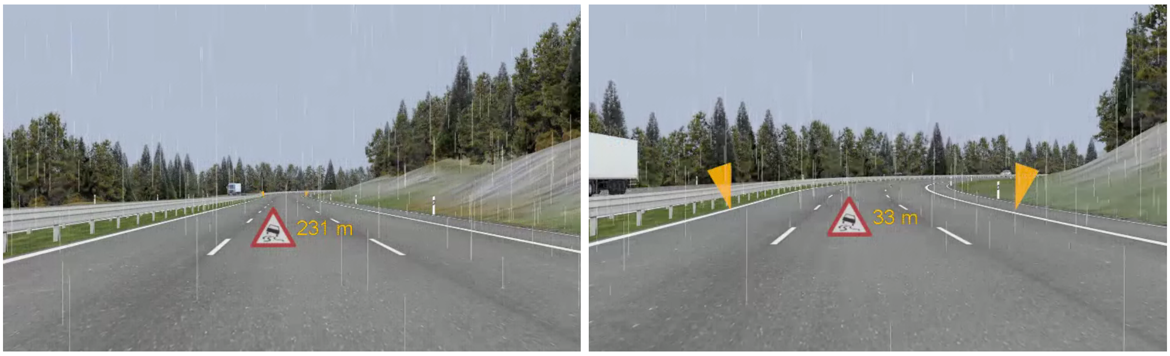


Figure 9.3: The narrow curve in heavy rain with the additional HMI at the moment of the RtI (left) and 33 meters away from the system limit (right).

(2016) showed that participants that are engaged in a visual NDRT would react to a RtI and direct their line of sight towards the road. If information such as the RtI was presented in the instrument cluster or the central information display (CID), participants would alternate their gazes between the road and these displays. Using a HUD, participants benefit from a reduced number of glances away from the road during the take-over process. In addition, Bubb et al. (2015, p. 412) recommended using the HUD to display information that could be directly related to the surroundings, e.g. the stopping distance, rather than showing vehicle-related internal information. In addition, a HUD reduces the number of glances away from the road and the reaction times in response to unexpected traffic incidents (Kiefer, 1998). Overall, the benefits of a HUD for displaying additional information during the take-over process are considerable.

The additional information consists of 2D parts that are stationary with respect to the HUD, like the traffic sign and the remaining distance. The system limit is highlighted using contact-analogue arrows on either side of the road. Such technology is currently researched and not available in current vehicles. Nevertheless, the contact-analogue arrows were a result of the underlying iterative design process and presented the best option to highlight the system limit. Several publications show the necessity to adequately inform drivers about system limits in general (Martens & van den Beukel, 2013), reasons for system behavior and resulting actions in critical situations (Koo et al., 2015) and the benefits from addressing psychological needs of drivers (Prasch & Tretter, 2016). The chosen information is based on these publications and the iterative design process. The HMI guidelines from the projects *UR:BAN* (Bengler et al., 2018, p. 75) and *AdaptIVe* (Kelsch et al., 2017) were considered in designing the HMI and were used as theoretical

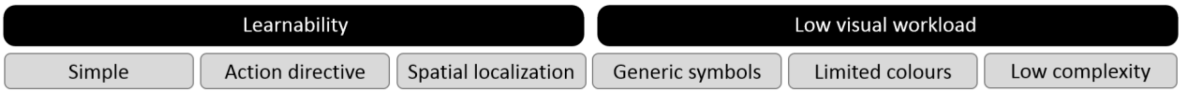


Figure 9.4: The general guidelines concerning visual displays for vehicles from the project *UR:BAN* (Bengler et al., 2018) were considered in designing the HMI. An expert review throughout the iterative design process was conducted along these guidelines.

Figure 9.5: The first and second draft of the proposed HMI that was discarded during the iterative design process.

basis for the design process. The guidelines from *UR:BAN* are shown in Figure 9.4. Concerning the iterative design process, a first draft suggested highlighting the surrounding vehicles, that are relevant for the take-over process in addition with providing a corridor in which the vehicle could be operated manually. However, a precise definition on which vehicles are relevant for the take-over process could not be reached. If an automation is advanced enough to provide semantic information on safe trajectories for manual drivers, it could be argued that such a system could solve the situation by itself. The first draft can be seen in Figure 9.5. The second draft already contained the display of the distance and the traffic sign. The system limit was visualized using a contact-analogue, semi-transparent bar that slightly hovered above the road, see 9.5. In a pretest with four HMI experts from the Chair of Ergonomics, it became apparent, that the bar would motivate braking to a full stop due to its resemblance with stop bars. Thus, the system limit was visualized using two arrows on either side of the road as can be seen in the final HMI.

The experimental setup, i.e. the automation, data logging and the eye-tracking system, was identical to the setup described in Chapter 4. Time budget was set to be seven seconds for both situations. The NDRT was the visual-motoric SuRT known from Experiments 1 and 2.

Measures

For data acquisition, the simulator with the simulation software SILAB allowed a tracking frequency of 120 Hz for all relevant vehicle dynamics. Both situations did not feature an obstacle and no additional traffic in the take-over, rendering an analysis of the TTC not practical. The *SmartEye* eye-tracking system with three cameras was used. For the analysis of the eye-tracking data, the ISO standard ISO/TS 15007 (2014) for filtering eye-tracking data which requires a 70 % detection rate was taken into account. In comparison with the detected areas of interest (AOIs) in SmartEye, data points which did not feature any recognized AOI were excluded from the data stream. Reasons for a non-recognition of any AOIs could be either due to closed eyes or the eye-tracker losing the current gaze. Since the experiment did not feature sleeping participants, the threshold-value of 70 %

would include many data points where the eye-tracker did not feature valid gaze tracking with a very high probability. In case the data stream per participant featured less than 70 % detected AOIs, affected participants were excluded from analysis. This led to group sizes of 17 for the group with HUD and 16 participants for the group with no HUD. After excluding participants from the analysis, which did not meet the 70 %-detection criteria, the overall mean of detected AOIs during the take-overs for all participants and both situations amounted to 89.9 %.

In this experiment, no seat pressure mats were used since the research focus was not on the development of the driver state during automated driving and its consequences on a take-over but rather on the benefit of an optimized HMI during the take-over. The HMI is analyzed using a set of questionnaires to allow evaluation of e.g. safety and usability aspects. Concerning the subjective rating of the situation, in addition to evaluating the subjective criticality, complexity and time budget, the obviousness of the reason for the RtI is analyzed. The usefulness and satisfaction of the HMI is rated with a customized questionnaire based on van der Laan, Heino, and de Waard (1997) that was translated and provided[2] by the Europe Chapter of the Human Factors and Ergonomics Society (HFES). The safety is assessed using five items, each to agree or disagree with on a five-point-Likert-scale, e.g. "The HMI can help lowering the crash risk" or "The HMI increases traffic safety." The safety questionnaire was adapted from Arndt (2011). The efficiency as part of usability is analyzed using two questions from Arndt (2011) concerning mistakes made by the HMI and if the information provided by the HMI is insufficient for drivers. Three questions based on Arndt (2011) focused on drivers' intentions of using the system in their vehicle, together with a question if drivers would buy such a system, also based on Arndt (2011). Three additional questions focused on drivers' willingness to engage in NDRTs during automated driving in general and if drivers wished to be informed on the reason for the RtI. The last question consisted of a multiple choice question asking which items of the additional HMI (arrows, traffic sign, distance, nothing additional and further items) were perceived.

Table 9.2 is providing an overview of the metrics that were recorded and analyzed.

Procedure

Participation in the experiment took a total of approximately 35 minutes and was rewarded with five euros. The experiment started with an introduction including the main focus of the experiment accompanied by a consent form. A demographic questionnaire captured relevant data for description of the sample after which participants were introduced to the simulator and the data acquisition including the eye-tracker. Participants were given a detailed instruction on the automation and the engagement into the NDRT. For the HMI group, the additional HMI was explained in the HUD. After calibration of the eye-tracker, participants could experience the vehicle dynamics in the simulator in manual driving mode, the automation and an exemplary take-over process and – for the group with additional HMI – the additional HMI during the take-over. The track of the experiment was 12.7 kilometers long and lasted approximately eight minutes. After each take-over situation, the questionnaires were presented via the intercom. After the final situation, participants were asked to park the vehicle in a rest area before the simulation was deactivated and participants answered the last questionnaire.

[2] http://www.hfes-europe.org/accept/accept_de.htm

Table 9.2: Summary of dependent variables used for assessing the take-over performance and the additional HMI.

Take-over performance and eye-tracking	Subjective ratings
Time aspects • Minimal time to the first deliberate action (steering wheel >2 ° or brake pedal >10 % or deactivation of automation by button) (take-over time, TOT) [s]	**Subjective rating of the situation** • Criticality of situation [ ] • Complexity of situation [ ] • Time budget of situation [ ] • Obviousness of reason for RtI [ ]
Quality aspects • Type of first reaction [ ] • Minimal (negative) and maximal longitudinal acceleration [m/s^2] • Maximal lateral acceleration (non-directional) [m/s^2] • Standard deviation of the lateral position (SDLP) [m]	**Subjective rating of the HMI** • Usefulness and satisfaction, semantic differential with nine pairs of adjectives [ ] • Perceived safety, five items [ ] • Efficiency as part of usability, two items [ ] • Intention to use the HMI in an own vehicle, four items [ ] • Willingness to engage in NDRTs [ ] • Wish to be informed on the reason for the RtI [ ]
Eye-Tracking • Percentage eyes on road (PEOR) [%] • Percentage eyes on Instrument Cluster (PEOIC) [%]	**Perceived elements of the HMI** • Perception of the items of the HMI, multiple choice question [ ]

9.3 Results

Analysis was conducted following the process depicted in Chapter 4.3. Concerning the values for longitudinal acceleration, a problem during data handling was revealed. The usual thresholds used for deactivation of the automation by steering (>2° steering wheel angle) or braking/accelerating (10 % pedal stroke) could be used to accurately analyze TOT and maximal lateral acceleration. Concerning the automation status, an internal simulation error did not deactivate the longitudinal component of the automation. Thus, values for the SDLP and the lateral acceleration consist of human input, whereas values for the longitudinal acceleration and velocity consist of superposed values from both automation and drivers. Since only few participants deactivated the automation by steering or pressing the button, the longitudinal part of the automation superposed by potential braking or accelerating inputs from drivers remained active during the take-over process for 77,5 % of participants. Consequently, the data and analysis for longitudinal acceleration was discarded. All other measures can be used for analysis of the take-over performance. Potential alterations of driver behavior during the take-over due to the parallel influence of manual and automation inputs can be argued. The superposed values of the longitudinal acceleration revealed mean values between -1 m/s^2 – 1 m/s^2 with very few outliers for the negative longitudinal acceleration reaching up to -8 m/s^2. Likely, participants which were braking hard can be identified from the data with the great majority of participants hardly braking at all. Both situations did not feature an obstacle or required braking prior to a potential lane change maneuver in order to reduce vehicle dynamics. While the superposed values for longitudinal acceleration cannot be used for analysis, the overall validity of the take-over performance data was assessed to be high.

9.3.1 Measures of take-over performance and eye-tracking results

Prior to the analysis of the TOT, the first reaction to the RtI (steering, braking or deactivating the automation using the button on the steering wheel) was assessed. Out of a total of 80 take-overs, nine participants did not take-over at all in the interval between RtI and system limit. In the construction site, the vehicle drifts out of lane since the situation does not feature any curve. The tight curve in heavy rain representing the other system limit starts right after the designated system limit. Participants not taking over in between the RtI and the system limit drift out of lane gradually. Even though the take-over in the familiarization drive consisted of a critical situation including an obstacle and the RtI was instructed clearly, nine participants did not take-over. After the system limit, they were asked to reactivate the automation in order to answer the questionnaires via intercom. The data from these nine participants were disregarded from analysis of take-over performance but not from analysis of eye-tracking data and the subjective ratings of both situation and subjective rating of the HMI. In case participants conducted a lane change maneuver in either take-over situation, data for the SDLP were discarded from analysis. Out of the total of 80 take-overs, in six cases a lane change was conducted.

Figure 9.6 shows the first type of reaction in response to the RtI for both groups and situations. Fisher's exact test does not reveal significant results ($p = .154$). Only few participants initially braked, with most participants initially steering or accelerating.

Analysis of results and prior tests concerning distribution and homogeneity of variance were conducted as depicted in Chapter 4.3. Since the within-factor only featured two levels, the requirement of sphericity was always met. Table C.1 shows an overview of the

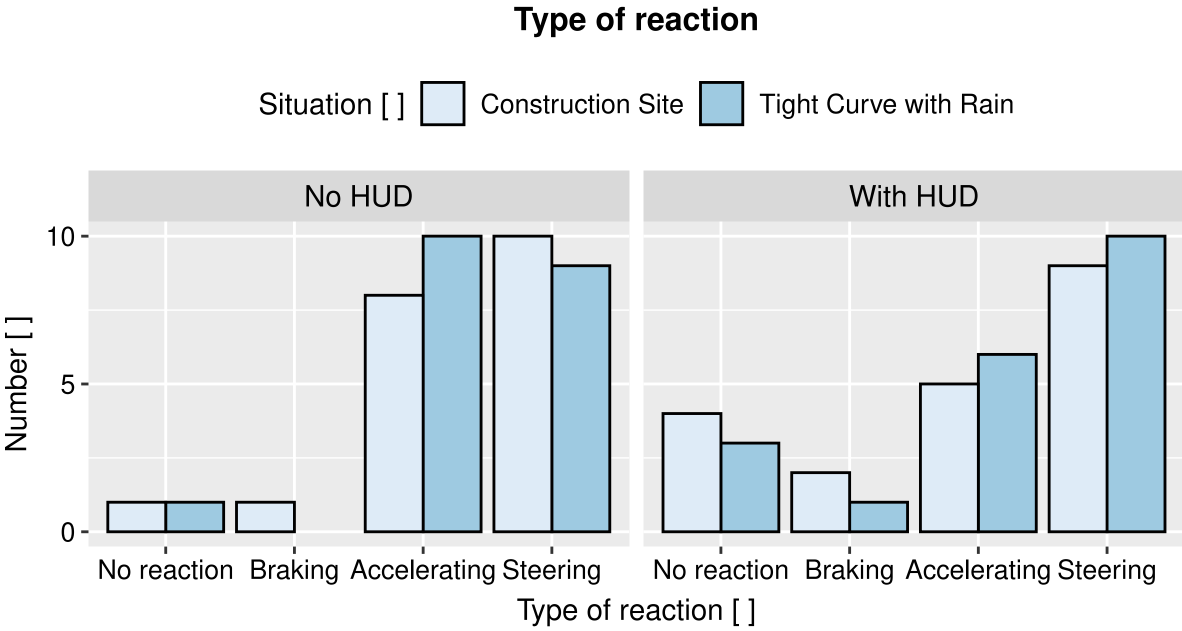


Figure 9.6: First reactions to the RtI for both groups and situations. $n_{\text{No HUD}}$ = 20, $n_{\text{With HUD}}$ = 20.

test results and the test statistics which are regarded for the evaluation of the research questions and the discussion of results. Data for the analysis of TOTs can be found in Figure 9.7. The ANOVA revealed no significant effects for the factor HUD, the factor Situation and the interaction between both groups (Table 9.4). For the analysis of the maximal lateral acceleration, the data were not differentiated between participants conducting a lane change or staying within their lane. In case a lane change was conducted, the maximal acceleration represents the absolute value. The ANOVA showed no significant effects for both factors and the interaction. Figure 9.8 shows the boxplot of the maximal lateral accelerations.

The standard deviation of lateral position was analyzed as measure of the quality of lateral control in the take-over. An ANOVA revealed no significant effects for both factors and their interaction (Table 9.4). The p-value for the factor HUD represents a tendency for the group with no additional HUD showing higher values for the SDLP. The data were visualized in Figure 9.9.

Concerning the AOIs for analysis, the percentage of eyes on road (PEOR) and the percentage of eyes on the instrument cluster (PEOIC) were regarded. For the PEOR, all glances through the windshield were regarded as "eyes on road". The ANOVA showed significant effects for the factor HUD, for the factor Situation and for the interaction HUD x Situation (Figure 9.10). The group with no additional HMI in the HUD showed significantly less PEOR during the take-over in the tight curve with heavy rain. The ANOVA for the PEOIC revealed a significant effect for the factor HUD but no significant effects for Situation and the interaction HUD x Situation (Figure 9.10). Results revealed that the group with no HUD had significantly more PEOIC compared to the group with HUD. Group means and SDs are listed in Table 9.3 while test statistics can be found in Table 9.4.

Table 9.3: Overview of the group and situation means (SD) for the objective take-over performance and eye-tracking data.

Dependent Variable	No HUD, construction site	No HUD, tight curve	HUD, construction site	HUD, tight curve
		TOT [s]		
TOT	M = 3.38 (1.62)	M = 3.23 (1.47)	M = 2.95 (1.13)	M = 3.64 (1.72)
		Maximal lateral acceleration [m/s^2]		
Lat. accel.	M = .73 (.51)	M = .85 (.56)	M = .58 (.65)	M = .86 (.86)
		SDLP [m]		
SDLP	M = .12 (.093)	M = .10 (.076)	M = .082 (.047)	M = .089 (.048)
		PEOR [%]		
PEOR	M = 74.76 (10.00)	M = 65.99 (9.90)	M = 77.94 (9.85)	M = 77.53 (7.55)
		PEOIC [%]		
PEOIC	M = 9.04 (7.23)	M = 12.76 (10.05)	M = 5.16 (5.49)	M = 5.67 (8.97)

Table 9.4: Results from the ANOVAs conducted for the objective take-over performance and eye-tracking measures.

Dependent Variable	Factor HUD, between groups	Factor Situation, within groups	Interaction HUD x situation
TOT	$F(1,31) = .00$, $p = .94, \eta^2 < .01$	$F(1,31) = 1.76$, $p = .19, \eta^2 = .02$	$F(1,31) = 1.70$, $p = .20, \eta^2 = .02$
Lateral acceleration	$F(1,38) = .47$, $p = .65, \eta^2 < .01$	$F(1,38) = .39$, $p = .16, \eta^2 = .02$	$F(1,38) = .39$, $p = .56, \eta^2 < .01$
SDLP	$F(1,32) = 3.31$, $p = .08, \eta^2 = .05$	$F(1,32) = .43$, $p = .52, \eta^2 < .01$	$F(1,32) = .12$, $p = .74, \eta^2 < .01$
PEOR	$F(1,31) = 6.88$, $p = .01, \eta^2 = .14$	$F(1,31) = 7.62$, $p = .01, \eta^2 = .06$	$F(1,31) = 6.34$, $p = .02, \eta^2 = .05$
PEOIC	$F(1,31) = 5.08$, $p = .03, \eta^2 = .11$	$F(1,31) = 2.17$, $p = .15, \eta^2 = .02$	$F(1,31) = 1.25$, $p = .27, \eta^2 = .01$

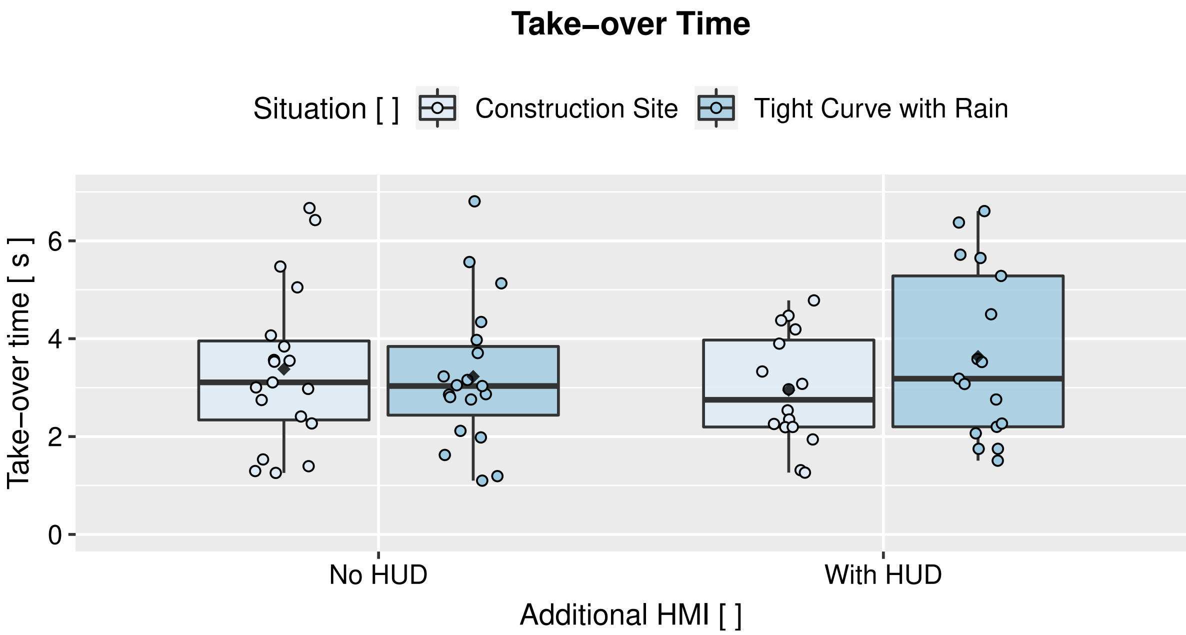


Figure 9.7: TOTs for both groups and situations. $n_{\text{No HUD}}$ = 38, $n_{\text{With HUD}}$ = 33.

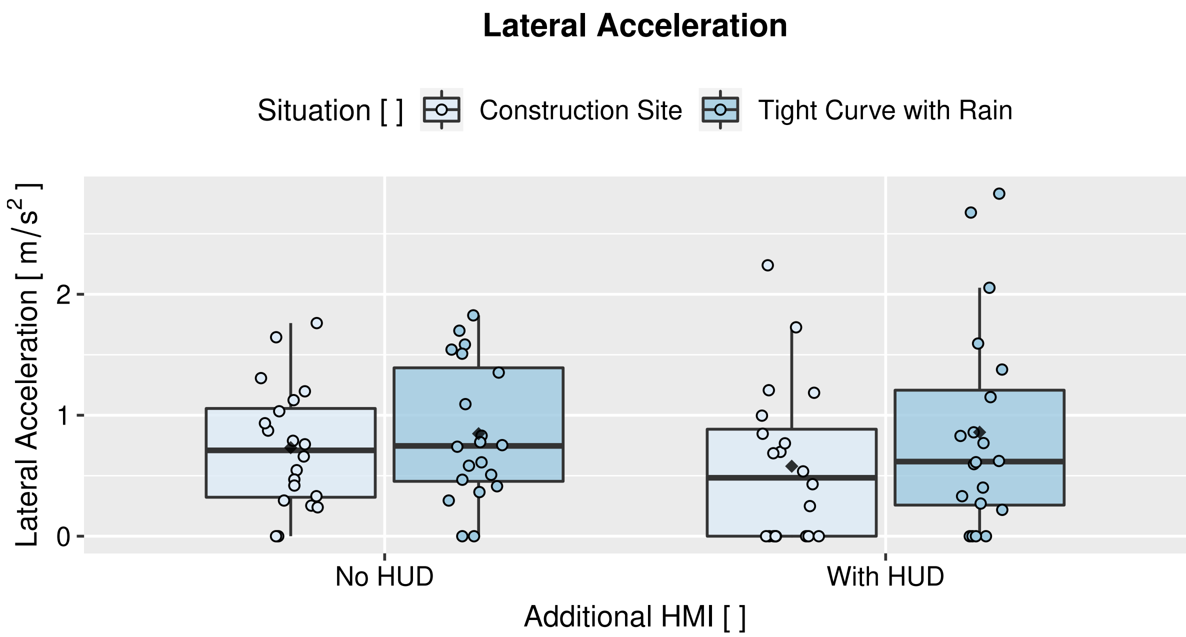


Figure 9.8: Maximal lateral acceleration for both groups and situations. $n_{\text{No HUD}}$ = 40, $n_{\text{With HUD}}$ = 40.

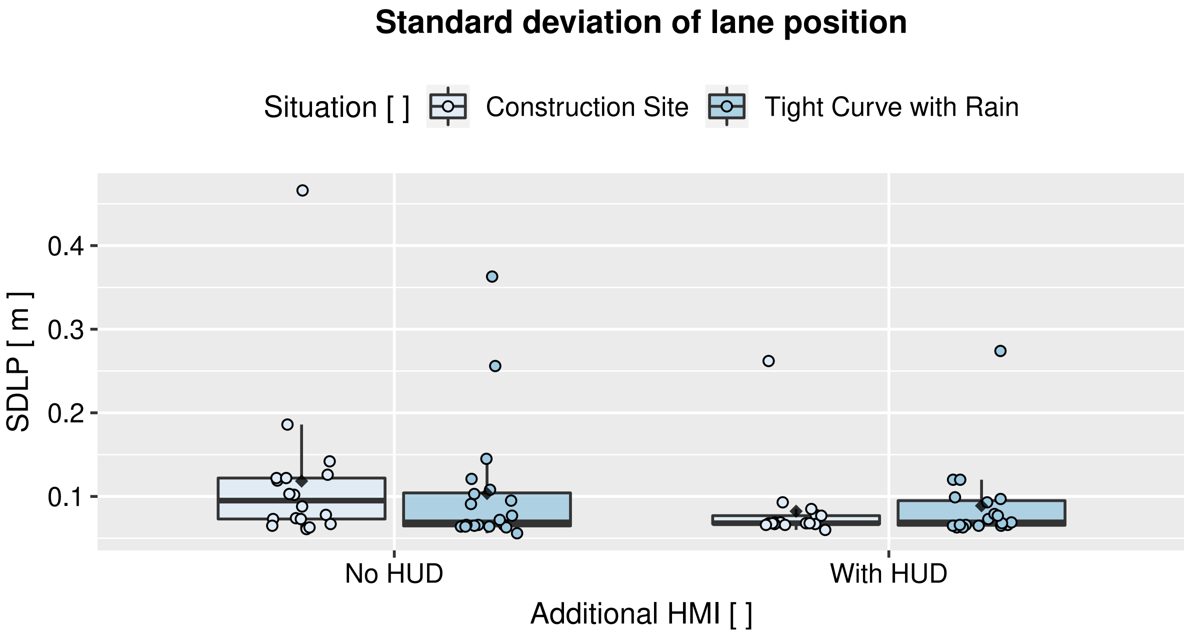


Figure 9.9: Maximal standard deviation of lane position (SDLP) for both groups and situations. $n_{\text{No HUD}} = 38$, $n_{\text{With HUD}} = 36$.

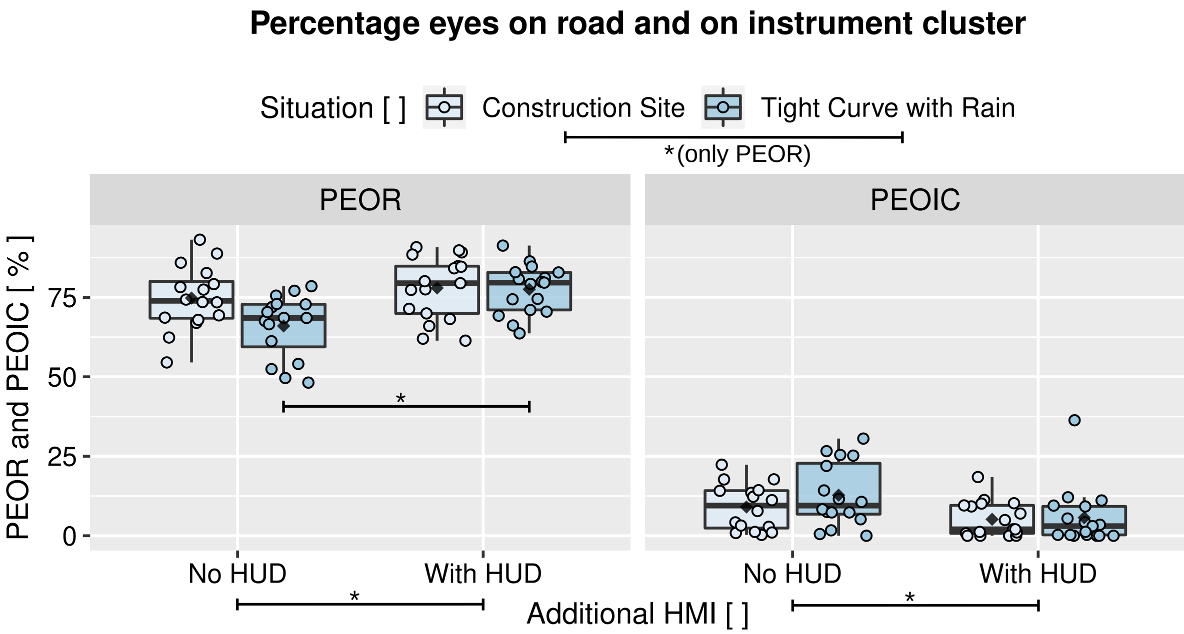


Figure 9.10: Percentage eyes on road (PEOR) and percentage eyes on instrument cluster (PEOIC) for both groups and situations. $n_{\text{No HUD}} = 32$, $n_{\text{With HUD}} = 34$.

9.3.2 Subjective ratings of the take-overs and the HMI

The analysis of the subjective ratings was conducted identically to the objective take-over performance and eye-tracking metrics. Data were checked for normal distribution and homogeneity of variance. Results can be found in Table C.2 and were regarded for the discussion of results. Table 9.5 shows the group means and SDs for the subjective ratings. Results for the perceived criticality showed that the construction site was assessed significantly more critical than the tight curve (Table 9.6). Values close to zero represent very uncritical ratings whereas values close to ten represent a critical subjective rating. The plotted data in Figure 9.11 show the significantly more critical ratings for the construction site. The factor HUD and the interaction between the groups and the situations did not show significant effects.

Analysis for the perceived complexity of the take-overs showed similar results with a significant effect for the factor Situation. The construction site was rated significantly more complex than the tight curve (Table 9.6). The factor HUD and the interaction did not show significant effects.

The perceived time budget was also analyzed with an ANOVA. Prior to analysis, the rating scale was turned around to allow a joined plot in Figure 9.11. Values close to zero represent participants felt that they had enough time in the take-over situation. Values close to ten would represent a subjectively high time pressure. The ANOVA showed significant effects for the factor Situation and the interaction between the groups and the situations (Table 9.6). The factor HUD showed no significant results. Participants rated the perceived time budget significantly lower (higher time pressure) for the construction site for the group with no HUD.

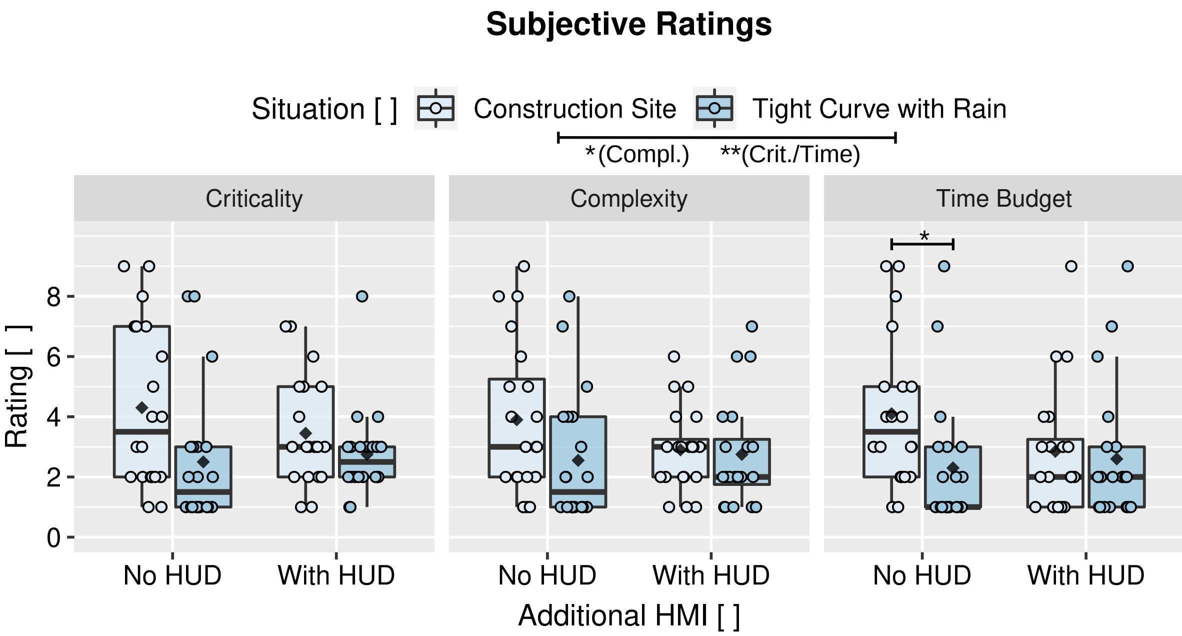


Figure 9.11: Subjective ratings of the perceived criticality, complexity and time budget for both groups and situations. $n_{\text{No HUD}} = 40$, $n_{\text{With HUD}} = 40$.

The obviousness of the reason for the take-over was ascertained after each situation. Results (Table 9.6) show that both the factor HUD and the factor Situation showed highly significant results. Referring to the group means (Table 9.5), the group with HUD reported a much higher obviousness of the reason for the take-over compared to the group with

Table 9.5: Overview of the group means (SDs) for the subjective ratings of the take-over situations.

Dependent Variable	No HUD, construction site	No HUD, tight curve	HUD, construction site	HUD, tight curve
		Criticality [ ]		
Criticality	M = 4.30 (2.72)	M = 2.50 (2.26)	M = 3.45 (1.82)	M = 2.75 (1.48)
		Complexity [ ]		
Complexity	M = 3.90 (2.53)	M = 2.55 (2.14)	M = 2.90 (1.37)	M = 2.75 (1.80)
		Time Budget [ ]		
Time Budget	M = 4.10 (2.47)	M = 2.30 (2.18)	M = 2.85 (2.11)	M = 2.60 (2.26)
		Obviousness [ ]		
Obviousness	M = 3.85 (1.42)	M = 2.00 (1.17)	M = 4.55 (0.69)	M = 3.55 (1.32)

Table 9.6: Results from the ANOVAs conducted for the subjective ratings.

Dependent Variable	Factor HUD, between groups	Factor Situation, within	Interaction HUD x situation
Criticality	$F(1,38) = .35$, $p = .56, \eta^2 < .01$	$F(1,38) = 8.21$, $p < .01, \eta^2 = .08$	$F(1,38) = 1.59$, $p = .22, \eta^2 = .02$
Complexity	$F(1,38) = .61$, $p = .44, \eta^2 = .01$	$F(1,38) = 4.05$, $p = .05, \eta^2 = .04$	$F(1,38) = 2.59$, $p = .12, \eta^2 = .02$
Time Budget	$F(1,38) = .55$, $p = .46, \eta^2 = .01$	$F(1,38) = 10.58$, $p < .01, \eta^2 = .05$	$F(1,38) = 6.05$, $p = .02, \eta^2 = .03$
Obviousness	$F(1,38) = 15.23$, $p < .001, \eta^2 = .19$	$F(1,38) = 35.66$, $p < .001, \eta^2 = .28$	$F(1,38) = 3.17$, $p = .08, \eta^2 = .03$

Table 9.7: Overview of the group and situation means (SDs) for the subjective ratings of the HMI in general after each situation.

Dependent Variable	No HUD, construction site	No HUD, tight curve	HUD, construction site	HUD, tight curve
		Usefulness [ ]		
Usefulness	M = -.15 (1.27)	M = -.25 (1.16)	M = 1.00 (.86)	M = .70 (.73)
		Satisfaction [ ]		
Satisfaction	M = -.20 (1.06)	M = -.55 (1.19)	M = .65 (.76)	M = .45 (.60)

no HUD. The tight curve showed significantly lower values for the obviousness. The interaction between the two factors showed a tendency for significant effects. Figure 9.12 underlines the results from the ANOVA. The HMI was also analyzed concerning

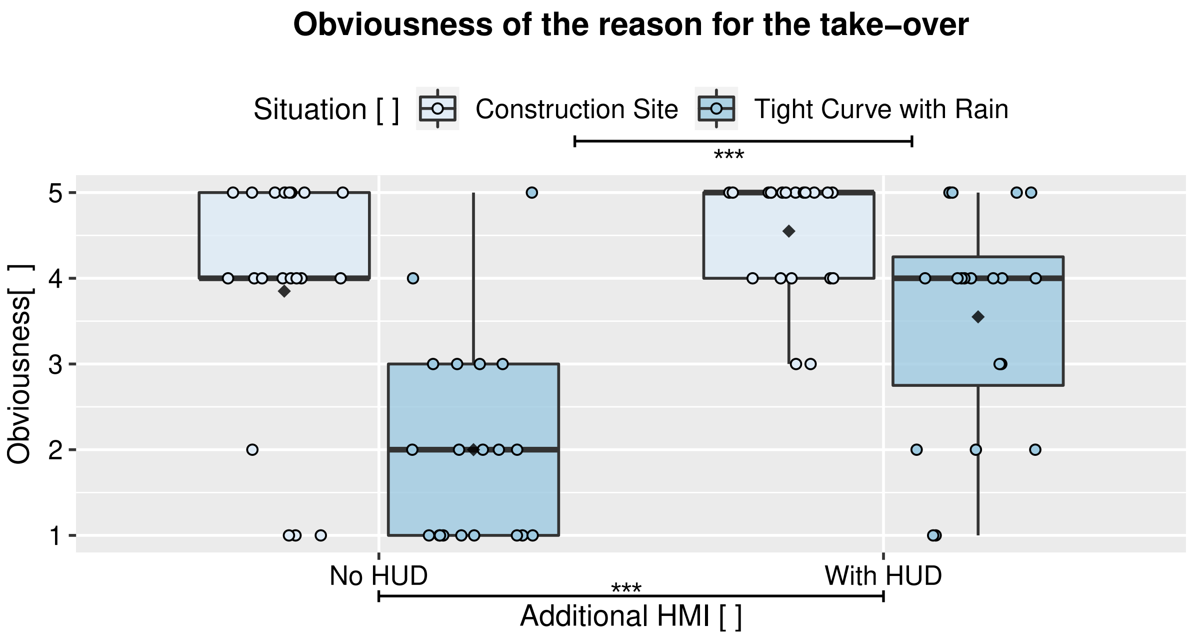


Figure 9.12: Obviousness of the reason for the take-over. $n_{\text{No HUD}} = 40$, $n_{\text{With HUD}} = 40$.

the subjective usefulness and satisfaction. An ANOVA revealed that the usefulness was rated significantly higher for the group with the additional HMI (Table 9.8). The factor Situation and the interaction between both factors showed no significant results. Results for satisfaction showed that the group with additional HUD rated the satisfaction significantly higher compared to the group with no HUD (Table 9.8) in addition to the construction site rated higher. Group means and SDs for both metrics can be found in Table 9.7 and are visualized in Figure 9.13.

After the experimental drive, participants filled out a final questionnaire on the overall rating of the HMI. The perceived safety (total of five items), the perceived efficiency as

Table 9.8: Results from the ANOVAs conducted for the subjective ratings of the HMI in general.

Dependent Variable	Factor HUD, between groups	Factor Situation, within groups	Interaction HUD x situation
Usefulness	$F(1,38) = 13.62$, $p < .001, \eta^2 < .22$	$F(1,38) = 1.60$, $p = .21, \eta^2 = .01$	$F(1,38) = .40$, $p = .53, \eta^2 = .002$
Satisfaction	$F(1,38) = 11.87$, $p = .001, \eta^2 < .21$	$F(1,38) = 5.29$, $p = .03, \eta^2 = .02$	$F(1,38) = .39$, $p = .53, \eta^2 = .002$

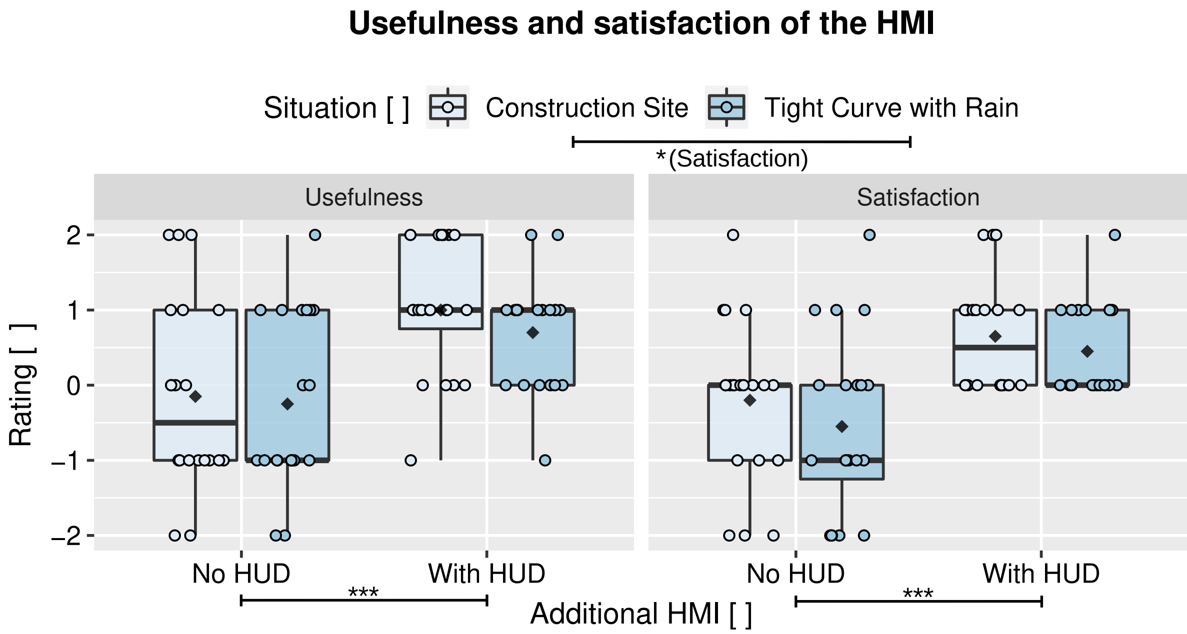


Figure 9.13: Subjective ratings of perceived usefulness and satisfaction of the HMI for both groups and situations. $n_{\text{No HUD}} = 40$, $n_{\text{With HUD}} = 40$.

part of the usability (total of two items) and the intention to use the HMI in an own vehicle (four items) were assessed on a five-point Likert-scale. Since the factor Situation was not regarded in the final questionnaire, only the factor HUD was analyzed using non-parametric tests in case of non-normal distributions. Table C.4 shows the test for normal distribution and homogeneity of variance for the three measures. Since the requirements for parametric tests were not met, Mann-Whitney-Wilcoxon tests (MWWT) were conducted. The results from the MWWTs showed highly significant effects for the perceived safety, the usability and the intention to use the HMI (Table 9.10). The plots in Figure 9.14 and the group means and SDs in Table 9.9 show that the group with the additional HUD reported significantly higher or better values for safety, usability and the intention to use.

As part of the final questionnaire, participants were asked if they could imagine engaging in NDRTs during CAD. In both groups, more than 75 % of participants answered the questions affirmative. The group with no additional HMI was asked if they would like to be informed about the reason for the take-over and more than 80 % of participants answered with "yes". The group with the additional HMI was asked which of the items in the additional HMI were perceived during the take-over. For all elements, the traffic

Table 9.9: Overview of the group means (SDs) for the subjective ratings of the HMI in the final questionnaire.

Dependent Variable	No HUD	HUD
Safety [ ]	M = 2.73 (.86)	M = 4.06 (.56)
Usability [ ]	M = 2.83 (1.17)	M = 3.90 (.72)
Intention to use [ ]	M = 2.25 (1.09)	M = 3.48 (.79)

Table 9.10: Results from the Mann-Whitney-Wilcoxon test (MWWT) conducted for the subjective ratings of the HMI after the experimental drive.

Dependent Variable	Factor HUD, between groups
Safety	$W = 43.5, p < .001, \eta^2 = .45$
Usability	$W = 93.0, p < .01, \eta^2 = .21$
Intention to use	$W = 76.0, p < .001, \eta^2 = .28$

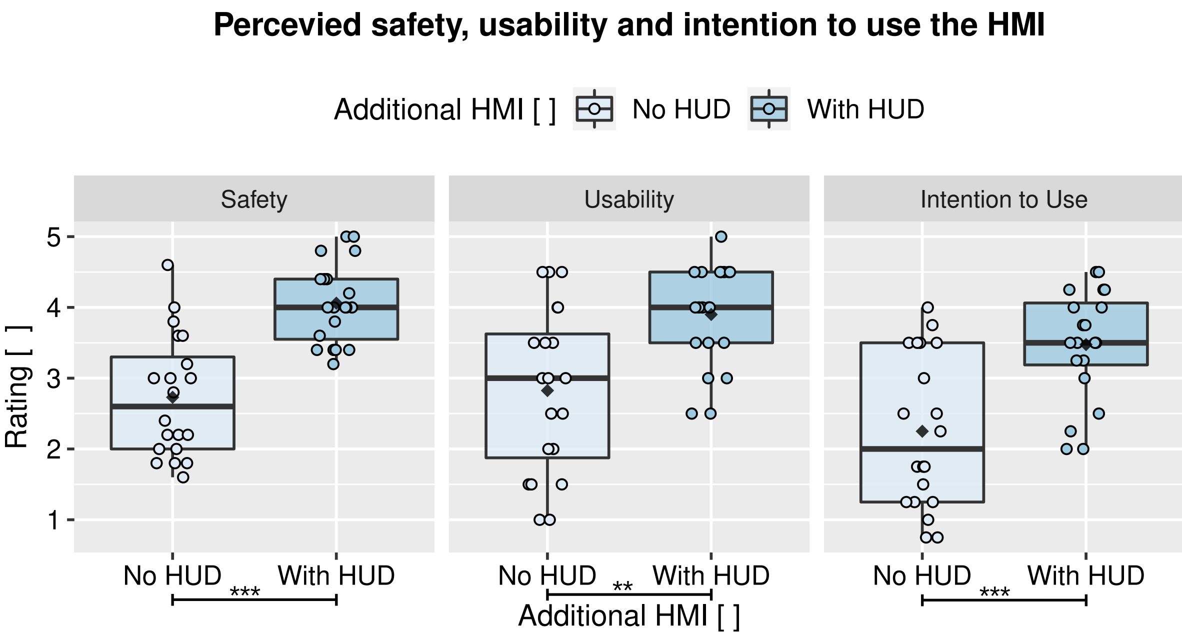


Figure 9.14: Subjective ratings of perceived safety, usability and the intention to use the HMI in an own vehicle. $n_{\text{No HUD}} = 20$, $n_{\text{With HUD}} = 20$.

sign representing the reason of the take-over, the distance towards the system limit and the arrows on each side of the road, more than 70 % of participants reported to have perceived all elements of the additional HMI. The traffic sign was perceived by 90 % of participants, representing the highest percentage of the three elements.

9.4 Discussion

Regarding the research questions in Figure 9.1, no significant differences were found for all measures of take-over performance. This can be attributed in part to the situation design, since both take-over situations consisted of an overall low criticality (Gold et al., 2017). While results from Experiments 1 and 2 and the modeling approach showed great situational differences, the findings in this chapter underline the prominent role of the situation in general: for overall non-critical situations, take-over performance is well within limits of known take-over performance values, showing no significant differences between similar situations concerning criticality.

In addition, take-over performance was unaffected by the introduction of the optimized HMI. While no significant improvement could be achieved, the additional information in the HUD did not overload participants during the take-over process. The eye-tracking measures showed significant improvements for the group with HMI, especially for the tight curve with rain. The significant interaction result for PEOR can be interpreted to show participants especially directing their visual attention towards the road in unclear take-over situations. The significant eye-tracking result did not affect the take-over performance, but including the significant contribution of PEOR for TOT and TTC in the modeling approach, adding visual information in the HUD is advised based on these findings. The eye-tracking behavior was only analyzed during the take-over, but gives way to solutions supporting drivers also before the take-over with semantic information in the HUD: in case participants choose to - temporarily - not direct their visual attention towards NDRTs, the optimized HMI could be beneficial.

Results also indicate that the manipulation of the obviousness was successful. Participants that encountered a take-over without an obvious reason for the RtI checked the instrument cluster more frequently. Results for the PEOIC for the group with no additional HUD showed significantly more glances towards the instrument cluster compared to the group with HUD. This indicates that participants are interested in understanding the reason for the take-over and in case of low obviousness would refer to the displays in the vehicle. If the information is presented in the instrument cluster or additional screens in the center console, the number of glances towards these displays increases.

Contrary to the results for the objective take-over performance, the subjective ratings of the take-over situations revealed significant situational differences. While TOT, lateral accelerations and SDLP indicate similar reactions in both situations, participants rated the construction site to be significantly more critical, complex, obvious and exerting higher time pressure, the latter only in the group with no additional HMI. A lack of reference of overall situational criticality in addition with an obvious reason for the RtI in the construction site could have led to these contradictory findings. A sense of not understanding the reason for the RtI together with no visible other objects or vehicles might have promoted the non-critical rating of the tight curve. The tendency of significant results for the interaction of obviousness shows that the HUD is likely to be beneficial in the less obvious situation tight curve and corroborates this conclusion. Overall, the additional information in the HUD sufficiently explains the reason for the RtI and is understood by participants. The interaction for the perceived time budget shows that the additional information in the HUD is actually leading to less perceived time pressure in the take-over even though participants are perceiving more information in general. Results indicate that the iterative design and development process resulted in an optimized HMI especially beneficial in situations that lack support due to the generic HMI.

Results for usefulness and satisfaction can be tied to the significantly lower obviousness of the situation: if participants had to take over but did not know the reason for it, the subjective satisfaction of the respective HMI could be lower because participants were missing substantial information for the less obvious situation. The analysis of the concluding questions after the experimental drive shows, that the additional information in the HUD was rated significantly better for all three aspects perceived safety, usability and intention to use revealing large effect sizes in comparison to small effects concerning the aforementioned subjective rating results. The findings are in line with literature presented in Chapter 2.2.4. The optimized HMI is more transparent regarding system limits and reasons for a RtI, satisfying general driver demands (Beggiato et al., 2015; J. Clark et al., 2017). The design of the take-over situations only showed effects in the subjective ratings, but a stronger effect of the optimized HMI inherently adapting to more critical situations can be hypothesized (Bazilinskyy et al., 2018). Any additional information should ideally be presented in the HUD since results from this work showed no deterioration of the take-over performance but significantly better ratings from drivers.

Concerning general limitations of this work, the data showed mostly non-normal distributions and the homogeneity of variance was violated in some cases. In combination with a skewed age distribution, the findings should not be generalized. A validation of results with older participants is recommended, since the technology of CAD will first be available in more expensive upper class vehicles, that are less likely purchased by younger drivers. Future work should include recently developed guidelines and verification work for HMIs designed for CAD (Naujoks, Wiedemann, et al., 2019), to avoid a wide raft of repeatedly optimized HMIs impeding a swift learning process throughout different platforms and commercial applications.

9.5 Summary and conclusion

Concluding, a generic HMI can be sufficient for perceiving a RtI but is lacking information necessary for not just a safe but also comfortable take-over process for participants. While the take-over performance was neither positively nor negatively affected by the additional information in the HUD, the subjective ratings from participants strongly benefited. In addition, results from eye-tracking showed that drivers with a HUD show less glances at in-vehicle displays, maximizing the time that could be utilized for visually perceiving the situation in a take-over. Most importantly, the findings from the subjective ratings highlight the necessity of sufficiently informing drivers about the reason for a take-over and aiding them in the take-over process. Drivers taking up their responsibility as fallback in case of a system limit are of utmost importance for the long-term success of CAD. Consequently, optimizing the take-over process should not only consider safety-related questions but focus on enhancing the overall experience of take-overs as well. This experiment provides empirical evidence that an optimization of the HMI by adding information in the HUD is perceived more useful, satisfactory, safer, usable and more likely to be appraised positively by drivers. The findings should be regarded for the ongoing research and development of HMIs for CAD.

10 General discussion

The individual results from Experiments 1-4 and the modeling approach were discussed in the individual chapters. This section discusses the overall results and limitations.

10.1 Limitations

The individual results are based on a well-established methodology concerning the design of experiments depicted in Chapter 4, but limitations apply regarding the validity and generalization of conclusions.

All results are based on samples of mainly German participants including few international students. Most conclusions potentially hold for Europe, but cultural differences were not regarded and should be addressed in future research since e.g. automotive HMI usability is influenced by cultural differences (Khan & Williams, 2014). In addition, traffic, past experiences and underlying traits likely differ between countries, potentially exerting effects on a variety of findings.

Most of the younger participants in all four experiments were students enrolled at the Technical University of Munich, typically at the Department of Mechanical Engineering. Absolute values of subjective ratings concerning the liking of CAD, the take-over or topics related to automated driving in general came from people with a higher technical affinity compared to the population and limits generalization of findings. The overall sample included a broader variety of participants as well, but findings presumably include a positive bias concerning CAD. This can also be understood to soften potential issues from experiencing newly developed technical system in addition with familiarization drives prior to every experimental drive. Moreover, the modeling approach from Gold (2016) is based on a similar group of participants. The comparison and validation of the models in Gold (2016) with results incorporating more heterogeneous samples showed reasonable prediction quality and give no cause to suspect low validity due to the sample in this work. Therefore, the lopsided distribution of the sample is not regarded to exert significant effects on the results in this thesis.

The experiments and the modeling approach allowed valuable insight with the empirical basis consisting of data gathered in the static simulator of the Chair of Ergonomics at the Technical University of Munich. The findings were not compared to real driving experiments and must be discussed critically. The Hawthorne-effect, giving reason to believe participants behave differently when knowing they are in an experiment, was shown to not affect the validity of driving simulator studies (Adair, 1984). Comparing manual driver reaction times in simulator and real driving environments showed statistical equivalence for important driver reaction metrics (McGehee, Mazzae, & Baldwin, 2000). More recent findings on the development of fatigue in manual driving, showed that fatigue can be equally studied in real and simulated driving environments (Philip et al., 2005). Latest research on comparing simulator with on-road findings in the context of automated driving indicates a high relative validity of driving simulators (Eriksson, Banks, & Stanton, 2017). Regarding the static nature of the employed driving simulator, absolute values of take-over quality, especially concerning accelerations and TTC must be regarded with care, while relative comparisons are dependable. TOT as measure of reaction time is regarded to be a valid measure concerning the transition from CAD to manual, but should be validated in on-road driving tests. Most importantly, Wizard-of-Oz studies found a comparable onset and development of drowsiness in combination with similar reaction times for simulated and real driving environments for CAD (Jarosch, Paradies, Feiner, & Bengler, 2019). The derivation of comparable take-over performance effects, at least concerning relative differences after both shorter and longer intervals of CAD seems reasonable. Concluding, analysis of effects from the individual experiments in addition with the relative comparison of different effects from the modeling approach is of high value for the overall conclusion of this thesis.

10.2 Levels of automated driving

The field of take-overs in CAD has seen a tremendous effort in past years to understand and design the inherent transitions between different levels of automation. The fundamental taxonomy used in this thesis and the current status quo in the field of research distinguishes between six levels of automated driving (SAE J3016, 2018).

Research on this taxonomy has yielded a critical view on the definition of CAD (Inagaki & Sheridan, 2018) and conflicting evidence that these six levels of automation do not necessarily meet the mental model of laymen (Homans, Radlmayr, & Bengler, 2019). The interaction between automated systems and drivers cardinally relies on their understanding, i.e. their mental model of the functions and capabilities of a particular system or level of automation (Sullivan, Flannagan, Pradhan, & Bao, 2016). Empirical work on drivers rating different functions concerning their level of automation showed that "in several instances, the functionality implied by automation terms did not match the technical definitions of the terms and/or the actual capabilities of the automated vehicle functions currently described by the terms" (Nees, 2018). This is in line with findings from Homans et al. (2019), which shed light on the fact that the levels of automation are feasible for the expert community but do not seem to represent the mental model of the majority of non-expert drivers. Contrary to the six levels of automated driving in the taxonomy, results showed a better representation of participants' mental model existing only of three levels. The differentiated distinction for Level 1 - assisted driving - to Level 4 - highly automated driving in the SAE taxonomy is merged into one level between manual driving and "robot vehicles" (Homans et al., 2019). A comparable study utilizing a different method for assessing the mental model of drivers also found a better depiction of levels of automated driving consisting of three levels and not six (Zacherl, Radlmayr, & Bengler, 2020). Other publications call for only two levels (driving vs. riding) (Seppelt et al., 2019), but the mismatch between the understanding of non-expert drivers and the expert taxonomy could harbor potential problems and should be addressed. While the overall dynamic of the topic of automated driving is calling for immediate answers, a more critical view on underlying assumptions is recommended. Human factors research on automated driving should be wary of evaluating and designing systems based on a taxonomy derived from a technical point of view but focus on actual user needs and expectations for vehicle automation.

Regarding a contraction of the various effects on take-over performance looking at situational effects only in perspective with the taxonomy, no limits concerning traffic situations are specified. The ODD allows a classification of functions and resulting responsibilities of drivers between levels of automation, but falls short of considering human factors issues of the take-over in CAD. Following the results from Experiments 1, 2 and the modeling approach, the fundamental conflict of interest concerning take-overs in CAD is highlighted again. While drivers show interest in CAD over PAD (Madigan, Louw,

& Merat, 2018) and the engagement in NDRTs, they are required to take-over in situations that cannot be managed by the automation but apparently play the most dominant role on the take-over performance themselves. Compared to the performance of manual drivers, more time and assistance might be required to reach the performance of manual driving in complex situations (Vogelpohl, Kühn, Hummel, Gehlert, & Vollrath, 2018). Taking long-term effects into account, participants could experience very few critical situations, adapting to take-overs being uncritical transitions and potentially losing the ability to be the fallback in critical situations. Or, participants could often be the fallback in critical situations, rather calling for the addition of minimal risk maneuvers (level 4) compared to CAD following the crash rate in Experiments 1 and 2.

A future framework or taxonomy including CAD as level of automated driving should integrate a critical view on situational effects to allow a successful introduction of CAD.

10.3 Summary

Research on take-over performance in the last years has provided a very valuable benchmark for findings and results of this work. The results from Experiments 1-4 and the modeling approach are generally in line with findings depicted in the literature overview in Chapter 2.

The TOT in CAD is highly influenced by the factor situation. Depending on the criticality and complexity of the specific situation, the TOT showed significant differences. More importantly, a period of prolonged automated diving - in this case 30 minutes - showed no significant changes concerning the TOT. All quality aspects of the take-over in addition with the subjective ratings were also unaffected comparing 5 and 30 minutes of automated driving without any NDRTs. Contrary, the highly significant effect of situation on take-over performance was also reported for accelerations, TTC and subjective ratings. Considering the measures of driver state comparing 5 and 30 minutes, significant differences were found for PERCLOS as measure of drowsiness and the changes of the COP in the seat and backrest. While appropriate measures reliably detect changes in the driver state, these do not result in different take-over performances. Regarding the results for NDRTs, a very similar conclusion can be drawn. "The influence of different criticality of the take-over situations is revealed and is consisted with findings from the overall scope of research. Eye-tracking and seat pressure mats offer a promising way of assessing changes in driver availability even though, in this experiment, they did not result in changes of the take-over performance accordingly" (Radlmayr, Fischer, & Bengler, 2019).

The joint modeling approach underlined the findings. TOT is highly affected by the factor situation, moderated by PEOR and age to a minor degree and significant amounts of variance were accounted for by the random effects structure. The highly significant influence of situation can be observed for the TTC and the accelerations alike, including the traffic density for TTC and lateral accelerations. In addition, PEOR influences the TOT and TTC, in part motivating the research questions of Experiments 3 and 4. Most importantly, the introduction of a random effects structure based on three or four take-overs per participants revealed differentiated findings concerning individual differences. TOT depends on the idiosyncratic, individual character of the drivers almost doubling the Pseudo-R^2 value of the model fit with/without random effects. In addition, the tendency to brake is highly dependent on individual participants, while the TTC and lateral accelerations show little to no effect from idiosyncratic effects in this work.

Future modeling of take-over performance should include random effects structures based on more take-overs to allow a better understanding of individual differences regarding limits of human performance for safety relevant take-overs. Aspects benefiting from more complex modeling approaches fostering a safety-relevant prediction quality include perceptual motor-skills (Mole et al., 2019) and guided transitions in case safety-relevant stages of re-entering manual control are missed (Vogelpohl & Vollrath, 2019).

Based on the significant results for PEOR and the conflict between engaging in NDRTs and visual attention being key to situation awareness, a possible solution was evaluated looking at peripheral monitoring. A visual NDRT was moved to the head-up display to allow peripheral monitoring of the surrounding traffic situation. While monitoring requests could increase the perception and understanding of the situation even during automated driving, they are in direct opposition of the definition of CAD where visual NDRTs are allowed. The possibility of peripheral monitoring was hypothesized to offer a potential solution to this conflict of interest. Results showed that situation awareness was degraded by engaging in a visual NDRT. The possibility of peripheral monitoring could not compensate this loss of situation awareness. Similar to Experiments 1 and 2, these differences in situation awareness did not lead to differences in take-over performance.

Regarding Experiment 4 on optimizing the HMI for a take-over, no differences in take-over performance were recorded again. Contrary, subjective ratings, such as usefulness, satisfaction, perceived safety and usability greatly benefited from introducing more information during the take-over compared to a generic HMI. In addition, the PEOR rate during the take-over was significantly higher for the group with the head-up display compared to the group with the generic HMI.

10.4 Future work and recommendations

Concluding, the overall findings suggest that take-over performance is highly depended on the specific situation at hand while state changes exert no or limited effects following the empirical basis of this work. Experiment 4 opened the door to recommending a stronger focus on the experience of a take-over for drivers in contrast to considering maximal human performance when drivers are the fallback. This would call for a stronger focus on e.g. interruption management to increase acceptance of automated driving (Naujoks, Wiedemann, & Schömig, 2017; Vogelpohl, Gehlmann, & Vollrath, 2019), adjusting trust in automation to avoid over-reliance (Radlmayr, Weinbeer, Löber, Farid, & Bengler, 2018; Körber, Baseler, & Bengler, 2018) or contemplating imperfect automation to reduce mental and physical demands (de Winter, Stanton, Price, & Mistry, 2016). The remaining challenge of drivers being fallback in situations too demanding for the automation must be addressed by various methods. Multi-stage concepts including pre-alerts have shown clear support of driver self-regulation and achieved high usability and acceptance ratings (Wandtner, 2018). The feedback of automation uncertainty and system states revealed an improvement of human-automation interaction (Beller, Heesen, & Vollrath, 2013) following fundamental guidelines from Norman (1990).

An application of CAD in real vehicles and the specific function characteristic should incorporate a situation recognition system. The technical challenges of detecting a system limit and issuing a RtI with time budgets relevant for feasible take-overs in CAD are highly demanding. Regardless, the dominant effect of the situation on the take-over performance in combination with findings from optimizing the HMI based on additional situational information calls for a function differentiating between situations. In case a situation reflects a system limit but is likely too critical and complex for drivers taking over, issuing a RtI seems unreasonable. While the implementation of compulsory minimal risk maneuvers denotes Level 4 systems, CAD systems should feature risk reducing maneuvers in highly critical situations. Overall, CAD features great benefits for drivers gaining time to engage in NDRTs of their choice in case the automation is available. The majority of take-overs is presumably in uncritical situations, allowing drivers to successfully assume their role as fallback. Nonetheless, the overlap between situations too critical for the automation and drivers must be regarded prior to the future application of CAD systems.

In addition, appropriate training of drivers concerning their changing and new role in automated vehicles provided benefits regarding performance optimal trust (Payre, Cestac, Dang, Vienne, & Delhomme, 2017; Forster, Hergeth, Naujoks, Beggiato, et al., 2019). Limitations concerning the underlying taxonomy and the definition of CAD can be addressed by a suitable training process. System-experienced drivers need to be considered for future application of CAD integrating long-term effects in the development process to avoid over-estimating first-exposure-effects (Stapel, Mullakkal-Babu, & Happee, 2019). The introduction of future technologies such as augmented reality in displays showed improvements in this work and in Langlois and Soualmi (2016). Underlying principle in this thesis was the deactivation of system support during a take-over to clearly analyze maximal human performance. Following the take-over process depicted in Figure 2.2, support of the driver during the take-over process is possible until manual control is firmly established. Potential applications should incorporate basic interaction principles for cooperative human-machine systems (Bengler, Zimmermann, Bortot, Kienle, & Damböck, 2012). Cooperative driving, understood as arbitrated, mutual control of the DDT by both the human and the automation has shown great potential concerning human factors' issues of transitioning between different levels of automation and manual driving (Fisher, Lohrenz, Moore, Nadler, & Pollard, 2016). While results suggest that drivers prefer shared control when an intervention is necessary (Mok, Johns, Gowda, Sibi, & Ju, 2016) and if they are primed timely (Kalb et al., 2018), the paradigm of cooperative control is not within the scope of this thesis.

Concluding the limitations and a critical discussion of findings, the consequences from an introduction of CAD on human performance need to be carefully considered to avoid seeing higher stages or levels of vehicle automation as universally beneficial, which is not the case (Onnasch, Wickens, Li, & Manzey, 2014). In contrast to high hopes in research and industry concerning the introduction of CAD, great caution should be exercised when assuming that drivers can "take-over" in case complex automated vehicles brake down (Kyriakidis et al., 2019).

10.5 Key messages

The thesis provides a thorough analysis of existing literature on CAD. Based on the derivation of research questions and their evaluation in four experiments and a modeling approach, effects on take-over performance regarding driver state changes and the optimization of the HMI were assessed and critically discussed. The seven most relevant key messages derived from this work are listed here.

This thesis offers the quantification of effects on take-over performance such as situational aspects and driver state. While these findings are embedded in a wide range of results from literature, for the first time, the individual, idiosyncratic contribution of drivers on measures such as TOT, TTC and accelerations is quantified as well. Based on the findings, the HMI for the take-over in CAD is optimized. Results call for a stronger focus on non-safety related aspects of human factors of CAD to foster a successful introduction in the near future.

1. State changes due to the paradigm change of CAD, engagement in NDRTs and longer periods of automated driving can be quantified using eye-tracking and seat pressure mats but have a minor or no effect on take-over performance.

2. Nonetheless, the driver state should be assessed in CAD to avoid precluded state changes like falling asleep and to detect preliminary stages of these such as high levels of drowsiness.

3. Time and quality aspects of take-over performance are highly depended on the specific situation in which the transition is experienced.

4. This is in line with the general consensus in the literature and the take-over situation can be considered to be of supreme importance in the assessment of safety and comfort of the take-over.

5. Individual differences between drivers exert substantial effects on TOT and braking behavior and should be regarded for the design and evaluation of CAD.

6. Due to the dominant effect of the take-over situation, drivers need to be supported by providing additional information on the situation at hand prior and during the transition process.

7. Focus should be put on maximizing the experience and subjective ratings of the take-over rather than addressing the limits of human performance which are inherently dependent on the specific situation.

Bibliography

Adair, J. G. (1984). The hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology*, *69*(2), 334-345. doi: 10.1037/0021-9010.69.2.334

Akaike, H. (1974). A new look at the statistical identification model. *IEEE Transactions on Automatic Control*, *19*, 716.

Arndt, S. (2011). *Evaluierung der Akzeptanz von Fahrerassistenzsystemen* (Dissertation). Technische Universität Dresden.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using r*. Cambridge University Press.

Baayen, R. H. (2009). *Correlation of fixed effects.* `https://stat.ethz.ch/pipermail/r-sig-mixed-models/2009q1/001941.html`. (Accessed: 2019-05-03)

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

Baldwin, C. L., & Coyne, J. T. (2003). Mental workload as a function of traffic density: Comparison of physiological, behavioral, and subjective indices. *Proceedings of the Second International Driving Symposium on Human Factors*, 19–24.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255 - 278. doi: https://doi.org/10.1016/j.jml.2012.11.001

Barua, S., Ahmed, M. U., Ahlström, C., & Begum, S. (2019). Automatic driver sleepiness detection using eeg, eog and contextual information. *Expert Systems with Applications*, *115*, 121 - 135. doi: https://doi.org/10.1016/j.eswa.2018.07.054

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Baumann, M., Rösler, D., & Krems, J. (2009). A procedure to estimate the effects of cognitive and visual distraction of in-car tasks on situation awareness while driving. In Gesellschaft für Arbeitswissenschaft (GfA) e.V. (Ed.), *Zeitschrift für Arbeitswissenschaft.* Ergonomia Verlag Stuttgart.

Bazilinskyy, P., Petermeijer, S., Petrovych, V., Dodou, D., & de Winter, J. (2018). Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays. *Transportation Research Part F: Traffic Psychology and Behaviour*, *56*, 82 - 98. doi: https://doi.org/10.1016/j.trf.2018.04.001

Beggiato, M., Hartwich, F., Schleinitz, K., Krems, J., Othersen, I., & Petermann-Stock, I. (2015). What would drivers like to know during automated driving? Information needs at different levels of automation. In *Proceedings of the 7th conference on driver assistance.*

Beller, J., Heesen, M., & Vollrath, M. (2013). Improving the Driver-Automation Interaction: An Approach Using Automation Uncertainty. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*(6), 1130–1141.

Bengler, K. (2015). Grundlegende Zusammenhänge von Automatisierung und Fahrerleistung. *Zeitschrift für Verkehrssicherheit*, *61*(3).

Bengler, K., Dietmayer, K., Farber, B., Maurer, M., Stiller, C., & Winner, H. (2014). Three decades of driver assistance systems: Review and future perspectives. *IEEE Intelligent Transportation Systems Magazine*, *6*(4), 6–22.

Bengler, K., Drüke, J., Hoffmann, S., Manstetten, D., & Neukum, A. (2018). *UR:BAN Human Factors in Traffic*. Wiesbaden: Springer Vieweg. doi: https://doi.org/10.1007/978-3-658-15418-9

Bengler, K., & Flemisch, F. (2011). Von H-Mode zur Kooperativen Fahrzeugführung: Grundlegende Ergonomische Fragestellungen. *5. Darmstädter Kolloquium - Zukunft der Fahrzeugführung. Kooperativ oder Autonom. Darmstadt, Germany*.

Bengler, K., Zimmermann, M., Bortot, D., Kienle, M., & Damböck, D. (2012). Interaction principles for cooperative human-machine systems. *it-Information Technology Methoden und innovative Anwendungen der Informatik und Informationstechnik*, *54*(4), 157-164.

Bentivoglio, A. R., Bressman, S. B., Cassetta, E., Carretta, D., Tonali, P., & Albanese, A. (1997). Analysis of blink rate patterns in normal subjects. *Movement Disorders*, *12*(6), 1028-1034. doi: 10.1002/mds.870120629

Bergasa, L. M., Nuevo, J., Sotelo, M. A., Barea, R., & Lopez, M. E. (2006). Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems*, *7*(1), 63-77.

Berghöfer, F. L., Purucker, C., Naujoks, F., Wiedemann, K., & Marberger, C. (2018). Prediction of take-over time demand in conditionally automated driving - Results of a real world driving study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting Europe Chapter*.

Borojeni, S. S., Chuang, L., Heuten, W., & Boll, S. (2016). Assisting Drivers with Ambient Take-Over Requests in Highly Automated Driving. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 237–244). New York, NY, USA: ACM. doi: 10.1145/3003715.3005409

Borojeni, S. S., Wallbaum, T., Heuten, W., & Boll, S. (2017). Comparing Shape-Changing and Vibro-Tactile Steering Wheels for Take-Over Requests in Highly Automated Driving. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 221–225). New York, NY, USA: ACM. doi: 10.1145/3122986.3123003

Borojeni, S. S., Weber, L., Heuten, W., & Boll, S. (2018). From Reading to Driving: Priming Mobile Users for Take-over Situations in Highly Automated Driving. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 14:1–14:12). New York, NY, USA: ACM. doi: 10.1145/3229434.3229464

Bortz, J., & Döring, N. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften (5., Aufl.)* [Book]. Berlin Heidelberg: Springer.

Bourrelly, A., Jacobé de Naurois, C., Zran, A., Rampillon, F., Vercher, J.-L., & Bourdin, C. (2019, August). Long automated driving phase affects take-over performance. *IET Intelligent Transport Systems*, *13*, 1249-1255(6).

Brandl, H. (2019). *Plot the equivalent of correlation matrix for factors (categorical data)? and mixed types?* https://stackoverflow.com/questions/52554336/plot-the-equivalent-of-correlation-matrix-for-factors-categorical-data-and-mi/56485520#56485520. (Accessed: 2020-04-07)

Braunagel, C., Rosenstiel, W., & Kasneci, E. (2017). Ready for Take-over? A New Driver Assistance System for an Automated Classification of Driver Take-Over Readiness. *IEEE Intelligent Transportation Systems Magazine*, *9*(4), 10-22. doi: 10.1109/MITS.2017.2743165

Brüch, K., Schmidt, K., Solbeck, C., & Wehner, T. (2017). *Periphere Verkehrsraumüberwachung beim hochautomatisierten Fahren (Scientific Thesis).* Chair of Ergonomics, Technical University of Munich, Munich, Germany.

Bubb, H., Bengler, K., Grünen, R. E., & Vollrath, M. (Eds.). (2015). *Automobilergonomie*. Wiesbaden: Springer Vieweg.

Bueno, M., Dogan, E., Selem, F. H., Monacelli, E., Boverie, S., & Guillaume, A. (2016). How different mental workload levels affect the take-over control after automated driving. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)* (p. 2040-2045). doi: 10.1109/ITSC.2016.7795886

Bühner, M., & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München, Harlow, Amsterdam, Madrid, Boston, San Francisco, Don MIlls, Mexiko City and Sydney: Pearson - Higher Education.

Cabrall, C., Janssen, N., Gonçalves, J., Morando, A., Sassman, M., & de Winter, J. (2016). Eye-based driver state monitor of distraction, drowsiness, and cognitive load for transitions of control in automated driving. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (p. 1981-1982). doi: 10.1109/SMC.2016.7844530

Card, S. K., Moran, T. P., & Newell, A. (1986). The Model Human Processor: An Engineering Model of Human Performance. *Handbook of Perception and Human Performance*, *2*, 1–45.

Chang, T.-H., & Chen, Y.-R. (2014). Driver fatigue surveillance via eye detection. *In 17th International IEEE Conference on Intelligent Transportation Systems, (ITSC), Qingdao, China*, 366–371.

Christensen, R. H. B. (2019). *ordinal – regression models for ordinal data.* (R package version 2019.12-10. https://CRAN.R-project.org/package=ordinal)

Clark, H., & 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 & Prevention*, *106*, 468 - 479. doi: https://doi.org/10.1016/j.aap.2016.08.027

Clark, J., Stanton, N. A., & Revell, K. (2017). Handover Assist Trials in Highly Automated Vehicles: Participant Recommendations for Future Design. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.*

Cohen-Lazry, G., & Katzman, N. (2018). Ipsilateral Versus Contralateral Tactile Alerts for Take-Over Requests in Highly-Automated Driving. In *6th Humanist Conference.*

Correlation between a nominal (iv) and a continuous (dv) variable. (2014). `https://stats.stackexchange.com/questions/119835/correlation-between-a-nominal-iv-and-a-continuous-dv-variable/124618#124618`. (Accessed: 2020-04-07)

Cramér, H. (1999). *Mathematical methods of statistics* (Vol. 43). Princeton University Press.

Damböck, D. (2013). *Automationseffekte im Fahrzeug – von der Reaktion zur Übernahme* (Dissertation). Technische Universität München.

Damböck, D., Farid, M., Tönert, L., & Bengler, K. (2012). Übernahmezeiten beim hochautomatisierten Fahren. *5. Tagung Fahrerassistenz. Munich, Germany*.

Darshana, S., Fernando, D., Jayawardena, S., Wickramanayake, S., & DeSilva, C. (2014). Efficient PERCLOS and Gaze Measurement Methodologies to Estimate Driver Attention in Real Time. In *2014 5th International Conference on Intelligent Systems, Modelling and Simulation* (p. 289-294). doi: 10.1109/ISMS.2014.56

David, B., Kuh, E., & Welsch, R. (1980). *Regression diagnostics: identifying influential data and sources of collinearity.* John Wiley & Sons, New York.

Daza, I. G., Hernandez, N., Bergasa, L. M., Parra, I., Yebes, J. J., Gavilan, M., ... Sotelo, M. A. (2011). Drowsiness monitoring based on driver and driving data fusion. *14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 1199–1204.

de Winter, J., Happee, R., Martens, M. H., & 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*, 196 - 217. (Vehicle Automation and Driver Behaviour) doi: https://doi.org/10.1016/j.trf.2014.06.016

de Winter, J., Stanton, N. A., Price, J., & Mistry, H. (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. doi: 10.1504/ijvd.2016.076736

Dinges, D., & Grace, R. (1998). PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance. *US Department of Transportation, Federal Highway Administration, Publication Number FHWA-MCRT-98-006*.

Dinges, D., & Mallis, M. (1998). Managing fatigue by drowsiness detection: Can technological promises be realized? In *International conference on fatigue and transportation, 3rd, 1998, fremantle, western australia.*

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46.

Endsley, M. (1988). Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *32*(2), 97–101.

Endsley, M. (1998). A comparative analysis of SAGAT and SART for evaluations of situation awareness. *Annual Meeting of the Human Factors & Ergonomics Society*, *42*.

Endsley, M. (2015a). Final Reflections: Situation Awareness Models and Measures. *Journal of Cognitive Engineering and Decision Making*, *9*(1), 101-111. doi: doi: 10.1177/1555343415573911

Endsley, M. (2015b). Situation Awareness Misconceptions and Misunderstandings. *Journal of Cognitive Engineering and Decision Making*, *9*(1), 4-32. doi: 10.1177/1555343415572631

Endsley, M., & Garland, D. (2000). *Direct Measurement of situation awareness: Validity and use of SAGAT.*

Epple, S., Roche, F., & Brandenburg, S. (2018). The Sooner the Better: Drivers' Reactions to Two-Step Take-Over Requests in Highly Automated Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 1883-1887. doi: 10.1177/1541931218621428

Eriksson, A., Banks, V., & Stanton, N. A. (2017). Transition to manual: Comparing simulator with on-road control transitions. *Accident Analysis & Prevention*, *102*, 227 - 234. doi: https://doi.org/10.1016/j.aap.2017.03.011

Eriksson, A., Petermeijer, S., Zimmermann, M., de Winter, J., Bengler, K., & Stanton, N. A. (2019). Rolling Out the Red (and Green) Carpet: Supporting Driver Decision Making in Automation-to-Manual Transitions. *IEEE Transactions on Human-Machine Systems*, *49*(1), 20-31. doi: 10.1109/THMS.2018.2883862

Eriksson, A., & Stanton, N. A. (2017). Driving Performance After Self-Regulated Control Transitions in Highly Automated Vehicles. *Human Factors*, *59*(8), 1233-1248. doi: 10.1177/0018720817728774

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, *77*, 167 - 181. doi: https://doi.org/10.1016/j.tra.2015.04.003

Feldhütter, A., Feierle, A., Kalb, L., & Bengler, K. (2018). A New Approach for a Real-Time Non-Invasive Fatigue Assessment System for Automated Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 1669-1673. doi: 10.1177/1541931218621379

Feldhütter, A., Kroll, D., & Bengler, K. (2018). Wake Up and Take Over! The Effect of Fatigue on the Take-over Performance in Conditionally Automated Driving. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (p. 2080-2085). doi: 10.1109/ITSC.2018.8569545

Feldhütter, A., Gold, C., Schneider, S., & Bengler, K. (2016). How Duration of Automated Driving Influences Take-Over Performance and Gaze Behavior. *62. GfA-Frühjahrskongress - Arbeit in komplexen Systemen - Digital, vernetzt, human?!*.

Feldhütter, A., Hecht, T., Kalb, L., & Bengler, K. (2018). Effect of prolonged periods of conditionally automated driving on the development of fatigue: with and without non-driving-related activities. *Cognition, Technology & Work*. doi: 10.1007/s10111-018-0524-9

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using r.* Sage publications.

Fischer, F. M. (2016). *Einfluss von fahrfremden Tätigkeiten auf die Fahrerverfügbarkeit beim hochautomatisierten Fahren (Master's Thesis).* Chair of Ergonomics, Technical University of Munich, Munich, Germany.

Fisher, D. L., Lohrenz, M., Moore, D., Nadler, E. D., & Pollard, J. K. (2016). Humans and Intelligent Vehicles: The Hope, the Help, and the Harm. *IEEE Transactions on Intelligent Vehicles*, *1*(1), 56-67. doi: 10.1109/TIV.2016.2555626

Flemisch, F., Bengler, K., Bubb, H., Winner, H., & Bruder, R. (2014). Towards cooperative guidance and control of highly automated vehicles: H-Mode and Conduct-by-Wire. *Ergonomics*, *57*(3), 343-360. doi: 10.1080/00140139.2013.869355

Flemisch, F., Nashashibi, F., Rauch, N., Schieben, A., Glaser, S., Temme, G., . . . Kaussner, A. (2010). Towards Highly Automated Driving: Intermediate report on the HAVEit-Joint System. *European Road Transport Research Arena*, *3*.

Forster, Y., Hergeth, S., Naujoks, F., Beggiato, M., Krems, J., & Keinath, A. (2019). Learning to use automation: Behavioral changes in interaction with automated driving systems. *Transportation Research Part F: Traffic Psychology and Behaviour*, *62*, 599 - 614. doi: https://doi.org/10.1016/j.trf.2019.02.013

Forster, Y., Hergeth, S., Naujoks, F., Krems, J., & Keinath, A. (2019). User education in automated driving: Owner's manual and interactive tutorial support mental model formation and human-automation interaction. *Information*, *10*(4). doi: 10.3390/info10040143

Forster, Y., Hergeth, S., Naujoks, F., Krems, J., & Keinath, A. (2020). Empirical validation of a checklist for heuristic evaluation of automated vehicle hmis. In N. A. Stanton (Ed.), *Advances in human factors of transportation* (pp. 3–14). Cham: Springer International Publishing.

Forster, Y., Naujoks, F., Neukum, A., & Huestegge, L. (2017). Driver compliance to take-over requests with different auditory outputs in conditional automation. *Accident Analysis & Prevention*, *109*, 18 - 28. doi: https://doi.org/10.1016/j.aap.2017.09.019

Franz, B., Haccius, J. P., Stelzig-Krombholz, D., Pfromm, M., Kauer, M., & Abendroth, B. (2015). Evaluation of the SAGAT method for highly automated driving. In *Proceedings 19th Triennial Congress of the IEA* (Vol. 9, p. 14).

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident Analysis & Prevention*, *37*(3), 461 - 472. doi: https://doi.org/10.1016/j.aap.2004.11.003

Gasser, T. M., & Westhoff, D. (2012). BASt-study: Definitions of automation and legal issues in Germany. In *Proceedings of the 2012 Road Vehicle Automation Workshop.*

Geiser, G. (1985). Mensch-Maschine-Kommunikation im Kraftfahrzeug. *ATZextra*, *87*(2), 77–84.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Gold, C. (2016). *Modeling of Take-Over Performance in Highly Automated Vehicle Guidance* (Dissertation). Technische Universität München, München.

Gold, C., Berisha, I., & Bengler, K. (2015). Utilization of Drivetime - Performing Non-Driving Related Tasks While Driving Highly Automated. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1666–1670.

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

Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking Over Control From Highly Automated Vehicles in Complex Traffic Situations: The Role of Traffic Density. *Human Factors: The Journal of the Human Factors and Ergonomics Society*.

Gold, C., Lorenz, L., & Bengler, K. (2014). Influence of Automated Brake Application on Take-Over Situations in Highly Automated Driving Scenarios. *Proceedings of FISITA World Automotive Congress. Maastricht, Netherlands*.

Gold, C., Naujoks, F., Radlmayr, J., Bellem, H., & Jarosch, O. (2017). Testing Scenarios for Human Factors Research in Level 3 Automated Vehicles. In N. A. Stanton (Ed.), *Advances in Human Aspects of Transportation: Proceedings of the AHFE 2017 International Conference on Human Factors in Transportation, Los Angeles, California, USA* (p. 551-559). Cham: Springer International Publishing. doi: 10.1007/978-3-319-60441-1_54

Gonçalves, J., & Bengler, K. (2015). Driver State Monitoring Systems– Transferable Knowledge Manual Driving to HAD. *Procedia Manufacturing*, *3*, 3011 - 3016. (6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015) doi: https://doi.org/10.1016/j.promfg.2015.07.845

Gonçalves, J., & Bengler, K. (2016). Challenges using portable devices during Conditional Automation. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)* (p. 2029-2033). doi: 10.1109/ITSC.2016.7795884

Gonçalves, J., Happee, R., & Bengler, K. (2016). Drowsiness in Conditional Automation: Proneness, diagnosis and driving performance effects. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)* (p. 873-878). doi: 10.1109/ITSC.2016.7795658

Gonçalves, R. C., Quaresma, M., & Rodrigues, C. M. (2017). Approaches for loss of vigilance in vehicle automation: a meta-analytical study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 1871-1875. doi: 10.1177/1541931213601948

Hafner, E. M. (2020). *"We're making autonomous driving safe." An interview about progress in development work.* https://www.daimler.com/innovation/case/autonomous/interview-hafner-2.html. (Accessed: 2020-01-03)

Happee, R., Gold, C., Radlmayr, J., Hergeth, S., & Bengler, K. (2017). Take-over performance in evasive manoeuvres. *Accident Analysis & Prevention*, *106*, 211 - 222. doi: https://doi.org/10.1016/j.aap.2017.04.017

Hecht, T., Feldhütter, A., Radlmayr, J., Nakano, Y., Miki, Y., Henle, C., & Bengler, K. (2019). A Review of Driver State Monitoring Systems in the Context of Automated Driving. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)* (pp. 398–408). Cham: Springer International Publishing.

Heikoop, D. D., de Winter, J., van Arem, B., & Stanton, N. A. (2016). Psychological constructs in driving automation: a consensus model and critical comment on construct proliferation. *Theoretical Issues in Ergonomics Science*, *17*(3), 284-303. doi: 10.1080/1463922X.2015.1101507

Helton, W. S., & Warm, J. S. (2008). Signal salience and the mindlessness theory of vigilance. *Acta Psychologica*, *129*(1), 18 - 25. doi: https://doi.org/10.1016/j.actpsy.2008.04.002

Hergeth, S., Lorenz, L., & Krems, J. (2017). Prior Familiarization With Takeover Requests Affects Drivers' Takeover Performance and Automation Trust. *Human Factors*, *59*(3), 457-470. doi: 10.1177/0018720816678714

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Homans, H., Radlmayr, J., & Bengler, K. (2019). Levels of driving automation from a user's perspective: How are the levels represented in the user's mental model? In *Proceedings of the international conference on human interaction & emerging technologies (IHIET).*

Inagaki, T., & Sheridan, T. B. (2018). A critique of the SAE conditional driving automation definition, and analyses of options for improvement. *Cognition, Technology & Work*. doi: 10.1007/s10111-018-0471-5

ISO/TS 14198. (2012). *Road vehicles - Ergonomic aspects of transport information and control systems - Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems* (Standard). Geneva, Switzerland: International Organization for Standardization, ISO.

ISO/TS 15007. (2014). *Measurement of driver visual behaviour with respect to transport information and control systems - Part 2: Equipment and procedures* (Standard). Geneva, Switzerland: International Organization for Standardization, ISO.

Ito, T., Takata, A., & Oosawa, K. (2016). Time Required for Take-over from Automated to Manual Driving. In *SAE Technical Paper Series.* SAE International. doi: 10.4271/2016-01-0158

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*(4), 394-412. doi: 10.1080/09658211003702171

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

Jarosch, O., Bellem, H., & Bengler, K. (2019). Effects of Task-Induced Fatigue in Prolonged Conditional Automated Driving. *Human Factors*, 0018720818816226. doi: 10.1177/0018720818816226

Jarosch, O., & Bengler, K. (2019). Is It the Duration of the Ride or the Non-driving Related Task? What Affects Take-Over Performance in Conditional Automated Driving? In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)* (pp. 512–523). Cham: Springer International Publishing.

Jarosch, O., Gold, C., Naujoks, F., Wandtner, B., Marberger, C., Weidl, G., & Schrauf, M. (2019). The impact of non-driving related tasks on take-over performance in conditionally automated driving – a review of the empirical evidence. In *9. Tagung Automatisiertes Fahren.* München.

Jarosch, O., Kuhnt, M., Paradies, S., & Bengler, K. (2017). It's Out of Our Hands Now! Effects of Non-Driving Related Tasks During Highly Automated Driving on Drivers' Fatigue. In *International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.*

Jarosch, O., Paradies, S., Feiner, D., & Bengler, K. (2019). Effects of non-driving related tasks in prolonged conditional automated driving – a wizard of oz on-road approach in real traffic environment. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 292 - 305. doi: https://doi.org/10.1016/j.trf.2019.07.023

Johns, M. (2000). A sleep physiologist's view of the drowsy driver. *Transportation Research Part F: Traffic Psychology and Behaviour*, *3*(4), 241-249. doi: https://doi.org/10.1016/S1369-8478(01)00008-0

Johns, M. (2007). Drowsy Driving and the Law. *Tasmania Law Reform Institute*, *1*(12).

Kaber, D. B., & Endsley, M. (1995). Out-of-the-Loop Performance Problems and the Use of Intermediate Levels of Automation for Improved Control System Functioning and Safety. *Human Factors*, *37*, 381–394.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs and N.J: Prentice-Hall.

Kalb, L., Streit, L., & Bengler, K. (2018). Multimodal Priming of Drivers for a Cooperative Take-Over. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (p. 1029-1034). doi: 10.1109/ITSC.2018.8569619

Kelsch, J., Dziennus, M., Schieben, A., Schömig, N., Wiedemann, K., Merat, N., ... Johansson, E. (2017). *AdaptIVe Deliverable D3.3 - Final functional Human Factors recommendations* (Report). `http://www.adaptive-ip.eu/index.php/deliverables_papers.html`: AdaptIVe Consortium.

Kerschbaum, P. (2017). *Design for Automation – The Steering Wheel in Highly Automated Cars* (Dissertation). Technische Universität München.

Kerschbaum, P., Lorenz, L., & Bengler, K. (2014). Highly automated driving with a decoupled steering wheel. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 1686–1690.

Kerschbaum, P., Lorenz, L., & Bengler, K. (2015). A Transforming Steering Wheel for Highly Automated Cars. *IEEE Intelligent Vehicles Symposium (IV)*, 1287–1292.

Khan, T., & Williams, M. (2014). A study of cultural influence in automotive hmi: Measuring correlation between culture and hmi usability. *SAE International Journal of Passenger Cars-Electronic and Electrical Systems*, *7*(2014-01-0263), 430–439.

Kiefer, R. J. (1998). Defining the HUD Benefit Time Window. *Vision in vehicles*, *6*, 133–142.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*, 352-358.

Ko, S. M., & Ji, Y. G. (2018). How we can measure the non-driving-task engagement in automated driving: Comparing flow experience and workload. *Applied Ergonomics*, *67*, 237 - 245. doi: https://doi.org/10.1016/j.apergo.2017.10.009

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2015). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *nternational Journal on Interactive Design and Manufacturing (IJIDeM)*, *9*(4), 269–275.

Kooperatives, hochautomatisiertes Fahren – Ko-HAF. (2018). `https://www.ko-haf.de/startseite/`. (Accessed: 2020-03-17)

Körber, M. (2018). *Inter-Individual Differences in Human-Automation Interaction - A Driver-Centered Perspective on the Introduction of Automated Vehicles* (Dissertation). Technische Universität München.

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, *66*, 18-31. doi: https://doi.org/10.1016/j.apergo.2017.07.006

Körber, M., Cingel, A., Zimmermann, M., & Bengler, K. (2015). Vigilance Decrement and Passive Fatigue Caused by Monotony in Automated Driving. *Procedia Manufacturing*, *3*, 2403-2409. doi: http://dx.doi.org/10.1016/j.promfg.2015.07.499

Körber, M., Gold, C., Lechner, D., & Bengler, K. (2016a). 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. doi: 10.1016/j.trf.2016.03.002

Körber, M., Gold, C., Lechner, D., & Bengler, K. (2016b). 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., Prasch, L., & Bengler, K. (2018). Why do i have to drive now? post hoc explanations of takeover requests. *Human Factors*, *60*(3), 305-323. doi: 10.1177/0018720817747730

Körber, M., Schneider, W., & Zimmermann, M. (2015). Vigilance, boredom proneness and detection time of a malfunction in partially automated driving. In *Collaboration technologies and systems (cts), 2015 international conference on* (p. 70-76). IEEE.

Kreuzmair, C., Gold, C., & Meyer, M.-L. (2017). The Influence of Driver Fatigue on Take-Over Performance in Highly Automated Vehicles. In *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration.*

Kyriakidis, M., de Winter, J., Stanton, N. A., Bellet, T., van Arem, B., Brookhuis, K., ... Happee, R. (2019). A human factors perspective on automated driving. *Theoretical Issues in Ergonomics Science*, *20*(3), 223-249. doi: 10.1080/1463922X.2017.1293187

Langenberg, J., Bartels, A., & Etemad, A. (2014). EU-Projekt AdaptIVe: Ansätze für hochautomatisches Fahren. *30. VDI/VW Gemeinschaftstagung Fahrerassistenz und Integrierte Sicherheit. Wolfsburg, Germany*, *30*.

Langlois, S., & Soualmi, B. (2016). Augmented reality versus classical hud to take over from automated driving: An aid to smooth reactions and to anticipate maneuvers. In *2016 IEEE 19th international conference on intelligent transportation systems (ITSC)* (p. 1571-1578). doi: 10.1109/ITSC.2016.7795767

Lorenz, L., Hergeth, S., Kerschbaum, P., Gold, C., & Radlmayr, J. (2015). Der Fahrer im Hochautomatisierten Fahrzeug. Vom Dual-Task zum Sequential-Task Paradigma: Ein Rückblick über Fahrsimulatorstudien. *7. Tagung Fahrerassistez. Munich, Germany*, *7*.

Lorenz, L., Kerschbaum, P., & Schumann, J. (2014). Designing take over scenarios for automated driving: How does augmented reality support the driver to get back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 1681–1685.

Louw, T. (2017). *The human factors of transitions in highly automated driving* (Dissertation). University of Leeds.

Louw, T., Merat, N., & Jamson, A. H. (2015). Engaging With Highly Automated Driving: To Be Or Not To Be In The Loop? *8th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design. Salt Lake City, USA*, *8*.

Lu, Z., Coster, X., & de Winter, J. (2017). How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Applied Ergonomics*, *60*, 293-304. doi: https://doi.org/10.1016/j.apergo.2016.12.003

Lu, Z., Happee, R., Cabrall, C. D. D., Kyriakidis, M., & de Winter, J. (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. doi: http://dx.doi.org/10.1016/j.trf.2016.10.007

Madigan, R., Louw, T., & Merat, N. (2018). The effect of varying levels of vehicle automation on drivers' lane changing behaviour. *PLOS ONE*, *13*(2), e0192190. doi: 10.1371/journal.pone.0192190

Manawadu, U. E., Hayashi, H., Ema, T., Kawano, T., Kamezaki, M., & Sugano, S. (2018, July). Tactical-level input with multimodal feedback for unscheduled takeover situations in human-centered automated vehicles. In *2018 IEEE/ASME international conference on advanced intelligent mechatronics (AIM)* (p. 634-639). doi: 10.1109/AIM.2018.8452227

Mangiafico, S. (2016). *cramerv: Cramer's v (phi).* `https://rdrr.io/cran/rcompanion/man/cramerV.html`. (Accessed: 2020-04-07)

Marberger, C., Mielenz, H., Naujoks, F., Radlmayr, J., Bengler, K., & Wandtner, B. (2017). Understanding and Applying the Concept of "Driver Availability" in Automated Driving. In N. A. Stanton (Ed.), *Advances in Human Aspects of Transportation: Proceedings of the AHFE 2017 International Conference on Human Factors in Transportation, Los Angeles, California, USA* (p. 595-605). Cham: Springer International Publishing. doi: 10.1007/978-3-319-60441-1_58

Martens, M. H., & van den Beukel, A. P. (2013). The road to automated driving: Dual mode and human factors considerations. *Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013)*, 2262–2267.

Mason, C. H., & Perreault, W. D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research*, *28*(3), 268.

Maurer, M., & Dickmanns, E. D. (1997). A system architecture for autonomous visual road vehicle guidance. In *Proceedings of conference on intelligent transportation systems* (p. 578-583). doi: 10.1109/ITSC.1997.660538

Maurer, M., Gerdes, J. C., Lenz, B., & Winner, H. (2016). *Autonomous driving* (Vol. 10).

May, J., & Baldwin, C. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*(3), 218-224.

Mazzega, J., & Schöner, H.-P. (2017). Wie PEGASUS die Lücke im Bereich Testen und Freigabe von automatisierten Fahrzeugen schließt. In *Methodenentwicklung für aktive Sicherheit und automatisiertes Fahren : 2. Expertendialog zu Wirksamkeit - Beherrschbarkeit - Absicherung* (pp. 163–175). Renningen: expert Verlag.

McCall, R., McGee, F., Mirnig, A., Meschtscherjakov, A., Louveton, N., Engel, T., & Tscheligi, M. (2018). A taxonomy of autonomous vehicle handover situations. *Transportation Research Part A: Policy and Practice*. doi: https://doi.org/10.1016/j.tra.2018.05.005

McDonald, A. D., Alambeigi, H., Engström, J., Markkula, G., Vogelpohl, T., Dunne, J., & Yuma, N. (2019). Toward computational simulations of behavior during automated driving takeovers: a review of the empirical and modeling literatures. *Human factors*, *61*(4), 642–688.

McGehee, D. V., Mazzae, E. N., & Baldwin, G. H. (2000). Driver Reaction Time in Crash Avoidance Research: Validation of a Driving Simulator Study on a Test Track. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(20), 3-320–3-323.

Melcher, V., Rauh, S., Diederichs, F., Widlroither, H., & Bauer, W. (2015). Take-Over Requests for Automated Driving. *Procedia Manufacturing*, *3*, 2867-2873. doi: http://dx.doi.org/10.1016/j.promfg.2015.07.788

Meteyard, L., & Davies, R. (2019, Mar). *Best practice guidance for linear mixed-effects models in psychological science.* PsyArXiv. Retrieved from `psyarxiv.com/h3duq` doi: 10.31234/osf.io/h3duq

Michon, J. A. (1985). A Critical View of Driver Behavior Models: What Do We Know, What Should We Do? In L. Evans & R. C. Schwing (Eds.), *Human Behavior and Traffic Safety* (pp. 485–524). Boston, MA: Springer US. doi: 10.1007/978-1-4613-2173-6_19

Miller, D., Sun, A., Johns, M., Ive, H., Sirkin, D., Aich, S., & Ju, W. (2015). Distraction Becomes Engagement in Automated Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1676-1680. doi: 10.1177/1541931215591362

Mok, B., Johns, M., Gowda, N., Sibi, S., & Ju, W. (2016). Take the wheel: Effects of available modalities on driver intervention. In *2016 IEEE Intelligent Vehicles Symposium (IV)* (p. 1358-1365). doi: 10.1109/IVS.2016.7535567

Mole, C. D., Lappi, O., Giles, O., Markkula, G., Mars, F., & Wilkie, R. M. (2019). Getting back into the loop: the perceptual-motor determinants of successful transitions out of automated driving. *Human factors*, *61*(7), 1037–1065.

Montgomery, D. C., & Peck, E. A. (1992). *Introduction to linear regression analysis* (2nd ed.). New York: Wiley.

Müller, C., & Bläsig, F. (2014). Die Fahrerbeobachtung als wichtiger Baustein für autonomes Fahren: Fahrerzustandsanalyse zur Steigerung der Sicherheit und Erweiterung des HMI. *Tagungsband 30. VDI/VW Gemeinschaftstagung*.

Murata, A., Koriyama, T., & Hayami, T. (2012). A basic study on the prevention of drowsy driving using the change of neck bending angle and the sitting pressure distribution. In *2012 Proceedings of SICE Annual Conference (SICE)* (p. 274-279).

Naujoks, F., Höfling, S., Purucker, C., & Zeeb, K. (2018). From partial and high automation to manual driving: Relationship between non-driving related tasks, drowsiness and take-over performance. *Accident Analysis & Prevention*, *121*, 28 - 42. doi: https://doi.org/10.1016/j.aap.2018.08.018

Naujoks, F., Purucker, C., & Neukum, A. (2017). A review of non-driving-related tasks used in studies on automated driving. In *8th Conference on Applied Human Factors and Ergonomics.*

Naujoks, F., Purucker, C., Wiedemann, K., & Marberger, C. (2019). Noncritical state transitions during conditionally automated driving on german freeways: Effects of non–driving related tasks on takeover time and takeover quality. *Human factors*, *61*(4), 596–613.

Naujoks, F., Wiedemann, K., & Schömig, N. (2017). The importance of interruption management for usefulness and acceptance of automated driving. In *ACM 9th*

International Conference on Automotive User Interfaces and Interactive Vehicular Applications.

Naujoks, F., Wiedemann, K., Schömig, N., Hergeth, S., & Keinath, A. (2019). Towards guidelines and verification methods for automated vehicle hmis. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 121 - 136. doi: https://doi.org/10.1016/j.trf.2018.10.012

Naujoks, F., Wiedemann, K., Schömig, N., Jarosch, O., & Gold, C. (2018). Expert-based controllability assessment of control transitions from automated to manual driving. *MethodsX*, *5*, 579 - 592. doi: https://doi.org/10.1016/j.mex.2018.05.007

Navarro, J. (2018). A state of science on highly automated driving. *Theoretical Issues in Ergonomics Science*, 1-31. doi: 10.1080/1463922X.2018.1439544

Nees, M. A. (2018). Drivers' Perceptions of Functionality Implied by Terms Used to Describe Automation in Vehicles. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 1893-1897. doi: 10.1177/1541931218621430

Nieuwenhuis, R., Te Grotenhuis, M., & Pelzer, B. (2012). influence.me: Tools for detecting influential data in mixed effects models. *R Journal*, *4*(2), 38-47.

Norman, D. A. (1990). The 'Problem' with Automation: Inappropriate Feedback and Interaction, not 'Over-Automation'. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *327*(1241), 585-593. doi: 10.1098/rstb.1990.0101

Omozik, K., Yang, Y., Kuntermann, I., Hergeth, S., & Bengler, K. (2019). How long does it take to relax? Observation of driver behavior during real-world conditionally automated driving. In *Proceedings of the Tenth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.*

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human Performance Consequences of Stages and Levels of Automation:An Integrated Meta-Analysis. *Human Factors*, *56*(3), 476-488. doi: 10.1177/0018720813501549

Parasuraman, R., & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*(3), 381–410.

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253.

Payre, W., Cestac, J., Dang, N.-T., Vienne, F., & Delhomme, P. (2017). Impact of training and in-vehicle task performance on manual control recovery in an automated car. *Transportation Research Part F: Traffic Psychology and Behaviour*, *46*, 216-227. doi: https://doi.org/10.1016/j.trf.2017.02.001

Petermeijer, S. (2017). *A vibrotactile interface to support the driver during the take-over process* (Dissertation). Technische Universität München, München.

Petermeijer, S., de Winter, J., & Bengler, K. (2015). Vibrotactile Displays: A Survey With a View on Highly Automated Driving. *IEEE Transactions on Intelligent Transportation Systems*, 1–11.

Pfleging, B., Rang, M., & Broy, N. (2016). Investigating user needs for non-driving-related activities during automated driving. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia* (p. 91-99). 3012735: ACM. doi: 10.1145/3012709.3012735

Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Åkerstedt, T., ... Bioulac, B. (2005). Fatigue, Sleepiness, and Performance in Simulated Versus Real Driving Conditions. *Sleep*, *28*(12), 1511-1516. doi: 10.1093/sleep/28.12.1511

Prasch, L., & Tretter, S. (2016). Ich bin dann mal raus. Erklärbares Automationsverhalten und Vertrauen. *Mensch und Computer 2016 – Kurzbeiträge, 4. - 7. September 2016, Aachen*.

Qiong, W., Jingyu, Y., Mingwu, R., & Yujie, Z. (2006). Driver Fatigue Detection: A Survey. In *6th World Congress on Intelligent Control and Automation* (Vol. 2, p. 8587-8591). doi: 10.1109/WCICA.2006.1713656

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Radlmayr, J., Brüch, K., Schmidt, K., Solbeck, C., & Wehner, T. (2018). Peripheral Monitoring of Traffic in Conditionally Automated Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 1828-1832. doi: 10.1177/1541931218621416

Radlmayr, J., Feldhütter, A., Frey, A., Jarosch, O., Marberger, C., Naujoks, F., ... Bengler, K. (2018). Drowsiness and fatigue in conditionally automated driving – Towards an integrative framework. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting Europe Chapter*.

Radlmayr, J., Fischer, F. M., & Bengler, K. (2019). The Influence of Non-driving Related Tasks on Driver Availability in the Context of Conditionally Automated Driving. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)* (pp. 295–304). Cham: Springer International Publishing.

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & 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 Annual Meeting*, *58*(1), 2063–2067.

Radlmayr, J., Ratter, M., Feldhütter, A., Körber, M., Prasch, L., Schmidtler, J., ... Bengler, K. (2019). Take-Overs in Level 3 Automated Driving – Proposal of the Take-Over Performance Score (TOPS). In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)* (pp. 436–446). Cham: Springer International Publishing.

Radlmayr, J., Weinbeer, V., Löber, C., Farid, M., & Bengler, K. (2018). How Automation Level and System Reliability Influence Driver Performance in a Cut-In Situation. In N. A. Stanton (Ed.), *Advances in Human Aspects of Transportation* (pp. 684–694). Cham: Springer International Publishing.

Ranney, T. A. (1994). Models of Driving Behavior: A Review of Their Evolution. *Accident Analysis & Prevention*, *26*(6), 733–750.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-13*(3), 257-266. doi: 10.1109/TSMC.1983.6313160

Rauch, N., Kaussner, A., Krüger, H.-P., Boverie, S., & Flemisch, F. (2009). The importance of driver state assessment within highly automated vehicles. *16th ITS World Congress. Stockholm, Schweden*, *21*, p. 25.

Roche, F., & Brandenburg, S. (2018). Should the urgency of auditory-tactile takeover requests match the criticality of takeover situations? In *21st International Conference on Intelligent Transportation Systems (ITSC)* (p. 1035-1040). doi: 10.1109/ITSC.2018.8569650

Roche, F., Somieski, A., & Brandenburg, S. (2019). Behavioral changes to repeated takeovers in highly automated driving: effects of the takeover-request design and the nondriving-related task modality. *Human factors*, *61*(5), 839–849.

SAE J3016. (2018). *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems* (Standard). Warrendale, PA: Society of Automotive Engineers, SAE.

Salkind, N. J. (Ed.). (2010). *Encyclopedia of research design*. Thousand Oaks, Calif.: SAGE Publ.

Scherer, L. (2016). *Modellierung der Fahrerverfügbarkeit beim hochautomatisierten Fahren (Master's Thesis).* Chair of Ergonomics, Technical University of Munich, Munich, Germany.

Schleicher, R., Galley, N., Briest, S., & Galley, L. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics*, *51*(7), 982–1010.

Schmidt, E., Decke, R., & Rasshofer, R. (2016). Correlation between subjective driver state measures and psychophysiological and vehicular data in simulated driving. In *2016 IEEE Intelligent Vehicles Symposium (IV)* (p. 1380-1385). doi: 10.1109/IVS.2016.7535570

Schmidt, E., Ochs, J., Decke, R., & Bullinger, A. (2017). Evaluating Drivers' States in Sleepiness Countermeasures Experiments Using Physiological and Eye Data – Hybrid Logistic and Linear Regression Model. In *Ninth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.* doi: 10.17077/drivingassessment.1648

Schmidt, J., Dreißig, M., Stolzmann, W., & Rötting, M. (2017). The Influence of Prolonged Conditionally Automated Driving on the Take-Over Ability of the Driver. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 1974-1978. doi: 10.1177/1541931213601972

Schmidt, J., Stolzmann, W., & Karrer-Gauß, K. (2016). Experimental evaluation of different request intervals for a driver alertness device for conditionally automated driving with induced drowsiness. *Proc. Automatisierungs-, Assistenzsysteme und eingebettete Systeme für Transportmittel*.

Schneider, J. H. (2009). *Modellierung und Erkennung von Fahrsituationen und Fahrmanövern für sicherheitsrelevante Fahrerassistenzsysteme* (Dissertation). Technische Universität Chemnitz.

Seppelt, B., Reimer, B., Russo, L., Bruce Mehler, B., Fisher, J., & Friedman, D. (2019, 06). Consumer confusion with with levels of vehicle automation. In *Proceedings of the tenth international driving symposium on human factors in driver assessment, training and vehicle design.*

Sheridan, T. B. (1966). Three models of preview control. *IEEE Transactions on Human Factors in Electronics*, *HFE-7*(2), 91-102. doi: 10.1109/THFE.1966.232329

Singmann, H. (2017). *Anova in r: afex may be the solution you are looking for.* `https://www.r-bloggers.com/anova-in-r-afex-may-be-the-solution-you-are-looking-for/`. (Accessed: 2019-03-16)

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2019). afex: Analysis of factorial experiments [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=afex` (R package version 0.23-0)

Singmann, H., & Kellen, D. (2017). An introduction to mixed models for experimental psychology. In *New methods in neuroscience and cognitive psychology.* Psychology Press Hove.

Sirkin, D., Martelaro, N., Johns, M., & Ju, W. (2017). Toward Measurement of Situation Awareness in Autonomous Vehicles. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (p. 405-415). ACM. doi: 10.1145/3025453.3025822

Skottke, E.-M., Debus, G., Wang, L., & Huestegge, L. (2014). Carryover Effects of Highly Automated Convoy Driving on Subsequent Manual Driving Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*(7), 1272–1283.

Stanton, N. A., & Young, M. S. (2000). A proposed psychological model of driving automation. *Theoretical Issues in Ergonomics Science*, *1*(4), 315-331. doi: 10.1080/14639220052399131

Stapel, J., Mullakkal-Babu, F. A., & Happee, R. (2019). Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 590 - 605. doi: https://doi.org/10.1016/j.trf.2018.11.006

Sullivan, J. M., Flannagan, M. J., Pradhan, A. K., & Bao, S. (2016). Literature review of behavioral adaptations to advanced driver assistance systems. *AAA Foundation for Traffic Safety*.

Teigen, K. H. (1994). Yerkes-Dodson: A Law for all Seasons. *Theory & Psychology*, *4*(4), 525–547.

Thiffault, P., & Bergeron, J. (2003). Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis & Prevention*, *35*(3), 381-391. doi: http://dx.doi.org/10.1016/S0001-4575(02)00014-3

Ulherr, A. D. (2019). *Bewertung des aktuellen Vorgehens bei Diskomfortuntersuchungen im Sitzen* (Dissertation). Technische Universität München, München.

van den Beukel, A. P., & van der Voort, M. C. (2013). The influence of time-criticality on Situation Awareness when retrieving human control after automated driving. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)* (p. 2000-2005). doi: 10.1109/ITSC.2013.6728523

van den Beukel, A. P., & van der Voort, M. C. (2014). Driver's situation awareness during supervision of automated control - comparison between SART and SAGAT measurement techniques. *European Conference on Human Centred Design for Intelligent Transport Systems. Vienna, Austria*, *4*.

van der Laan, J., Heino, A., & de Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, *5*(1), 1–10. doi: 10.1016/S0968-090X(96)00025-3

Vanhove, J. (2019). *Collinearity isn't a disease that needs curing.* `https://janhove.github.io/analysis/2019/09/11/collinearity`. (Accessed: 2020-04-07)

Vlakveld, W., van Nes, N., de Bruin, J., Vissers, L., & van der Kroft, M. (2018). Situation awareness increases when drivers have more time to take over the wheel in a level 3 automated car: A simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 917 - 929. doi: https://doi.org/10.1016/j.trf.2018.07.025

Vogelpohl, T., Gehlmann, F., & Vollrath, M. (2019). Task interruption and control recovery strategies after take-over requests emphasize need for measures of situation awareness. *Human factors*, 0018720819866976.

Vogelpohl, T., Kühn, M., Hummel, T., Gehlert, T., & Vollrath, M. (2018). Transitioning to manual driving requires additional time after automation deactivation. *Transportation Research Part F: Traffic Psychology and Behaviour*, *55*, 464-482. doi: https://doi.org/10.1016/j.trf.2018.03.019

Vogelpohl, T., Kühn, M., Hummel, T., & Vollrath, M. (2018). Asleep at the automated wheel – sleepiness and fatigue during highly automated driving. *Accident Analysis & Prevention*. doi: https://doi.org/10.1016/j.aap2018.03.013

Vogelpohl, T., & Vollrath, M. (2019). Attentional dynamics after take-over requests: The need for handover assistance systems in highly automated vehicles. In H. Krömker (Ed.), *HCI in Mobility, Transport, and Automotive Systems* (pp. 128–142). Cham: Springer International Publishing.

Walch, M., Mühl, K., Kraus, J., Stoll, T., Baumann, M., & Weber, M. (2017). From Car-Driver-Handovers to Cooperative Interfaces: Visions for Driver–Vehicle Interaction in Automated Driving. In G. Meixner & C. Müller (Eds.), *Automotive User Interfaces: Creating Interactive Experiences in the Car* (p. 273-294). Cham: Springer International Publishing. doi: 10.1007/978-3-319-49448-7_10

Wan, J., & Wu, C. (2018). The effects of lead time of take-over request and nondriving tasks on taking-over control of automated vehicles. *IEEE Transactions on Human-Machine Systems*, *48*(6), 582-591. doi: 10.1109/THMS.2018.2844251

Wandtner, B. (2018). *Non-driving related tasks in highly automated driving - Effects of task characteristics and drivers' self-regulation on take-over performance* (Dissertation). Universität Würzburg.

Wandtner, B., Schömig, N., & Schmidt, G. (2018). Effects of non-driving related task modalities on takeover performance in highly automated driving. *Human factors*, *60*(6), 870–881.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance Requires Hard Mental Work and Is Stressful. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 433–441.

Weinbeer, V., Baur, C., Radlmayr, J., Bill, J.-S., Muhr, T., & Bengler, K. (2017). Highly automated driving: How to get the driver drowsy and how does drowsiness influence various take-over-aspects? In Lehrstuhl für Fahrzeugtechnik mit TÜV SÜD Akademie (Ed.), *8. Tagung Fahrerassistenz.*

Weinbeer, V., Bill, J.-S., Baur, C., & Bengler, K. (2018). Automated driving: subjective assessment of different strategies to manage drowsiness. *Proceedings of the human factors and ergonomics society europe*.

Weinbeer, V., Muhr, T., & Bengler, K. (2019). Automated Driving: The Potential of Non-driving-Related Tasks to Manage Driver Drowsiness. In *20th Congress of the International Ergonomics Association (IEA 2018)* (p. 179-188). Springer International Publishing.

Werner, S. (2018). *Konzeption und Evaluation eines Human Machine Interfaces für eine optimierte übernahmequalität und -zeit beim hochautomatisierten Fahren durch periphere Blickführung mithilfe des Head-up-Displays (Master's Thesis).* Chair of Ergonomics, Technical University of Munich, Munich, Germany.

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455.

Wickens, C. D., & Liu, Y. (1988). Codes and Modalities in Multiple Resources: A Success and a Qualification. *Human Factors*, *30*(5), 599-616. doi: 10.1177/001872088803000505

Wickens, C. D., & McCarley, J. S. (2007). *Applied attention theory.* CRC press.

Wierwille, W. W., & Ellsworth, L. A. (1994). Evaluation of driver drowsiness by trained raters. *Accident Analysis & Prevention*, *26*(5), 571 - 581. doi: https://doi.org/10.1016/0001-4575(94)90019-1

Williams, M., & Preston, N. C. (1987). The prometheus project. *6th International Conference on Automotive Electronics*, *6*(280), 36–39.

Winkler, S., Kazazi, J., & Vollrath, M. (2018). How to warn drivers in various safety-critical situations – Different strategies, different reactions. *Accident Analysis & Prevention*. doi: https://doi.org/10.1016/j.aap.2018.01.040

Winner, H., Hakuli, S., Lotz, F., & Singer, C. (2015). *Handbuch Fahrerassistenzsysteme: Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort* (3., berarb. und erg. Aufl. ed.). Wiesbaden: Springer Vieweg.

Winter, B. (2013). A very basic tutorial for performing linear mixed effects analyses. *arXiv preprint arXiv:1308.5499*.

Winter, B. (2019). *Statistics for linguists: An introduction using r*. Taylor & Francis. Retrieved from https://books.google.de/books?id=tnm6DwAAQBAJ

Wright, T. J., Samuel, S., Borowsky, A., Zilberstein, S., & 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. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 270-273. doi: 10.1177/1541931213601062

Yang, Y., Karakaya, B., Dominioni, G. C., Kawabe, K., & Bengler, K. (2018). An HMI Concept to Improve Driver's Visual Behavior and Situation Awareness in Automated Vehicle. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (p. 650-655). doi: 10.1109/ITSC.2018.8569986

Yang, Y., Klinkner, J. N., & Bengler, K. (2019). How Will the Driver Sit in an Automated Vehicle? – The Qualitative and Quantitative Descriptions of Non-Driving Postures (NDPs) When Non-Driving-Related-Tasks (NDRTs) Are Conducted. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)* (p. 409-420). Springer International Publishing.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology and psychology*, *18*(5), 459–482.

Zacherl, L., Radlmayr, J., & Bengler, K. (2020). Constructing a mental model of automation levels in the area of vehicle guidance. In *International conference on intelligent human systems integration* (pp. 73–79).

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

Zeeb, K., Buchner, A., & 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 & Prevention*, *92*, 230-239.

Zhang, B., de Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *64*, 285 - 307. doi: https://doi.org/10.1016/j.trf.2019.04.020

Zilberg, E., Xu, Z. M., Burton, D., Karrar, M., & Lal, S. (2009). Statistical validation of physiological indicators for non-invasive and hybrid driver drowsiness detection system. *African Journal of Information & Communication Technology*, *5*(2). doi: 10.5130/ajict.v5i2.1128

Zimmermann, M., & Bengler, K. (2013). A multimodal interaction concept for cooperative driving. *IEEE Intelligent Vehicles Symposium (IV)*, 1285–1290.

Zuur, A. F., & Ieno, E. N. (2016). *Beginner's guide to zero-inflated models with r*. Highland Statistics Ltd.

A Appendix for Chapter 5 "The effects of prolonged conditionally automated driving on driver state and take-over performance"

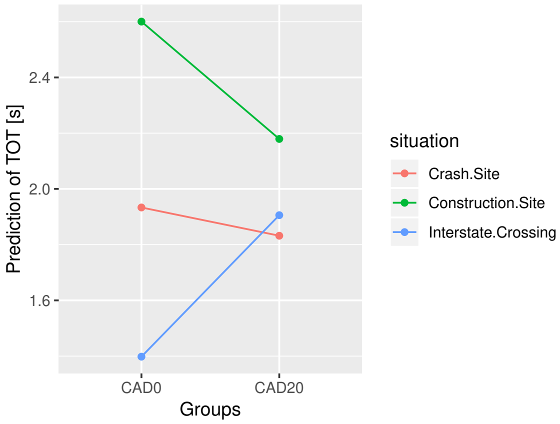


Figure A.1: Plot of the point estimates for TOT revealing the significant effect for traffic density and situation and their interaction.

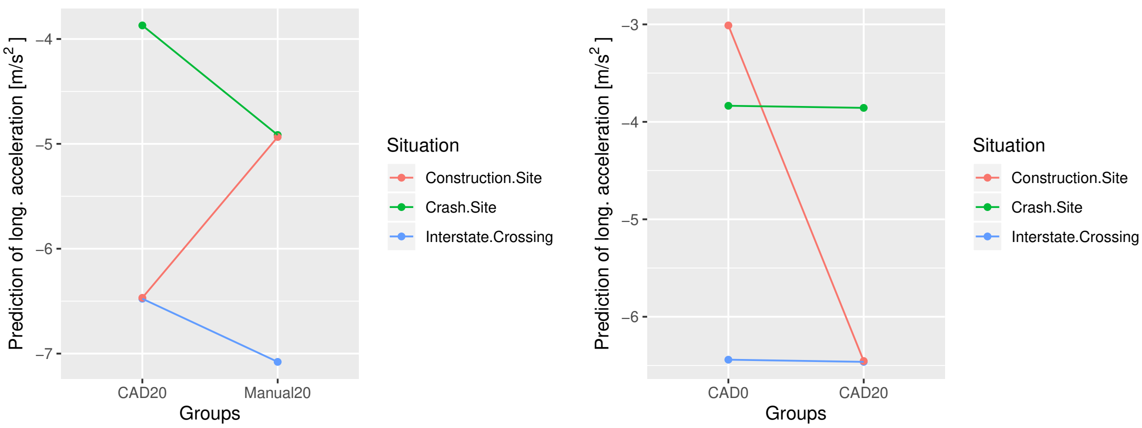


Figure A.2: Plot of the point estimates for the longitudinal acceleration revealing the significant effect for situation and the interaction effects.

Table A.1: Overview of test results on normal distribution and homogeneity of variance for the TOT.

Measure	Shapiro-Wilk	Levene	F_{max}
	Take-over time		
Groups	CAD0: $W = .859, p < .001,$ CAD20: $W = .962, p = .07$	$F(1, 115) = .19,$ $p = .67$	-
Situations	Crash: $W = .960, p = .06,$ Construction: $W = .892, p < .001,$ Crossing: $W = .954, p = .03$	- (within)	-
Durations	5A: $W = .903, p < .001,$ 5B: $W = .904, p < .001,$ 30: $W = .908, p < .001$	- (within)	-

Table A.2: Overview of test results on normal distribution and homogeneity of variance for the longitudinal accelerations.

Measure	Shapiro-Wilk	Levene	F_{max}
	Longitudinal acceleration		
Groups	CAD0: $W = .783, p < .001$, CAD20: $W = .792, p < .001$, Manual: $W = .816, p < .001$	$F(2, 167) =$ $.55, p = .58$	-
Situations	Crash: $W = .810, p < .001,$ Construction: $W = .777, p < .001,$ Crossing: $W = .714, p < .001$	- (within)	-
Durations	5A: $W = .822, p < .001$, 5B: $W = .803, p < .001$, 30: $W = .795, p < .001$	- (within)	-

Table A.3: Overview of test results on normal distribution and homogeneity of variance for the lateral accelerations.

Measure	Shapiro-Wilk	Levene	F_{max}
	Lateral acceleration		
Groups	CAD0: $W = .956, p = .03$, CAD20: $W = .942, p < .01$, Manual: $W = .921, p < .01$	$F(2, 167) =$ $5.92, p < .01$	2.2
Situations	Crash: $W = .939, p < .01$, Construction: $W = .837, p < .001$, Crossing: $W = .781, p < .001$	- (within)	-
Durations	5A: $W = .951, p = .02$, 5B: $W = .952, p = .03$, 30: $W = .907, p < .001$	- (within)	-

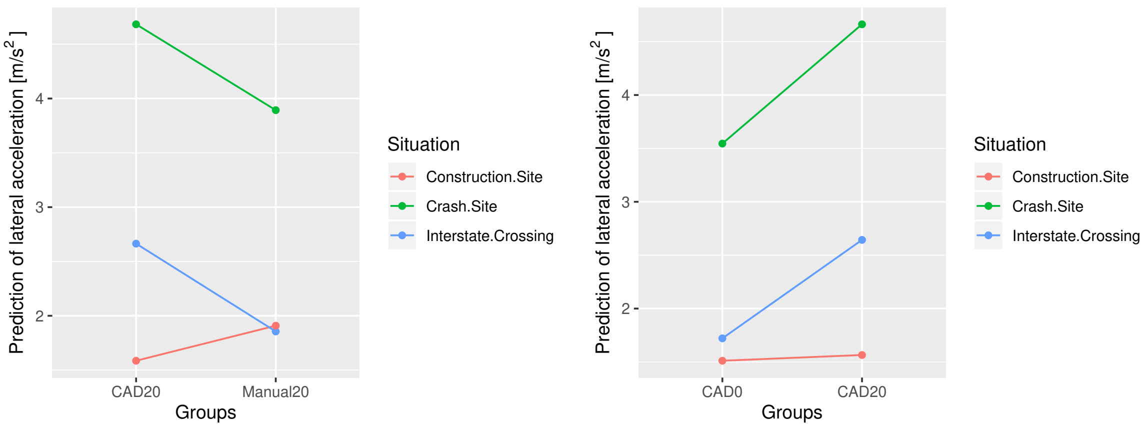


Figure A.3: Plot of the point estimates for the lateral acceleration revealing the significant effect for situation and the factor traffic density in the right plot.

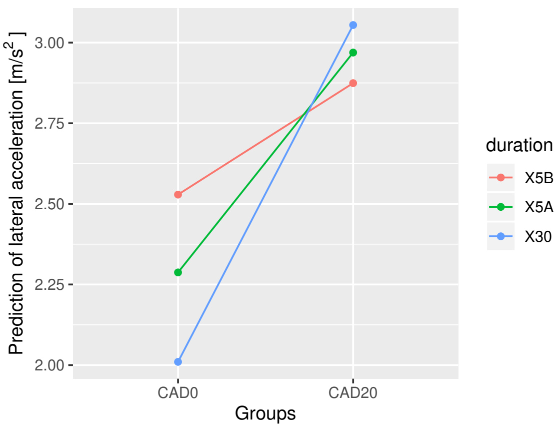


Figure A.4: Plot of the point estimates for the lateral acceleration revealing the significant effect for traffic density.

Table A.4: Overview of test results on normal distribution and homogeneity of variance for the TTC.

Measure	Shapiro-Wilk	Levene	F_{max}
	Lateral acceleration		
Groups	CAD0: $W = .972, p = .18$, CAD20: $W = .906, p < .001$, Manual: $W = .899, p < .01$	$F(2,148) = .71, p = .49$	-
Situations	Crash: $W = .926, p = .002$, Construction: $W = .942, p = .05$, Crossing: $W = .962, p = .08$	- (within)	-
Durations	5A: $W = .935, p = .008$, 5B: $W = .928, p = .005$, 30: $W = .964, p = .14$	- (within)	-

Table A.5: Overview of test results on normal distribution and homogeneity of variance for the subjective ratings.

Measure	Shapiro-Wilk	Levene	F_{max}
	Subjective criticality		
Groups	CAD0: $W = .947, p = .01$, CAD20: $W = .957, p = .04$, Manual: $W = .931, p < .01$	$F(2, 168) = .43, p = .65$	-
Situations	Crash: $W = .924, p < .01$, Construction: $W = .943, p < .01$, Crossing: $W = .937, p < .01$	- (within)	-
	Subjective complexity		
Groups	CAD0: $W = .955, p = .03$, CAD20: $W = .973, p = .23$, Manual: $W = .967, p = .14$	$F(2, 168) = 2.04, p = .13$	-
Situations	Crash: $W = .968, p = .14$, Construction: $W = .948, p = .02$, Crossing: $W = .962, p = .07$	- (within)	-
	Subjective comfort		
Groups	CAD0: $W = .919, p < .001$, CAD20: $W = .928, p < .01$, Manual: -	$F(1, 115) = 1.12, p = .29$	-
Situations	Crash: $W = .934, p = .02$, Construction: $W = .923, p = .01$, Crossing: $W = .924, p = .01$	- (within)	-
	Subjective time budget		
Groups	CAD0: $W = .920, p < .001$, CAD20: $W = .949, p = .02$, Manual: $W = .916, p = .001$	$F(2, 167) = .60, p = .55$	-
Situations	Crash: $W = .953, p = .03$, Construction: $W = .924, p = .002$, Crossing: $W = .915, p < .001$	- (within)	-

Table A.6: Overview of test results on normal distribution and homogeneity of variance for the eye-tracking measures during automated driving.

Measure	Shapiro-Wilk	Levene	F_{max}
	HGD		
Groups	CAD0: $W = .954, p = .04$, CAD20: $W = .986, p = .79$	$F(1, 106) = .07, p = .79$	-
Duration	5A: $W = .959, p = .18$, 5B: $W = .954, p = .15$, 30: $W = .975, p = .58$	- (within)	-
	PEOR		
Groups	CAD0: $W = .906, p < .001$, CAD20: $W = .780, p < .001$	$F(1, 106) = .20, p = .65$	-
Duration	5A: $W = .887, p = .001$, 5B: $W = .773, p < .001$, 30: $W = .870, p < .001$	- (within)	-
	PERCLOS		
Groups	CAD0: $W = .512, p < .001$, CAD20: $W = .533, p < .001$	$F(1, 104) = .65, p = .42$	-
Duration	5A: $W = .565, p < .001$, 5B: $W = .513, p < .001$, 30: $W = .592, p < .001$	- (within)	-
	Blink duration		
Groups	CAD0: $W = .880, p < .001$, CAD20: $W = .927, p < .01$	$F(1, 105) = .01, p = .93$	-
Duration	5A: $W = .879, p < .001$, 5B: $W = .948, p = .10$, 30: $W = .872, p < .001$	- (within)	-
	Blink frequency		
Groups	CAD0: $W = .868, p < .001$, CAD20: $W = .917, p < .01$	$F(1, 105) = 1.48, p = .23$	-
Duration	5A: $W = .881, p < .001$, 5B: $W = .844, p < .001$, 30: $W = .897, p < .01$	- (within)	-

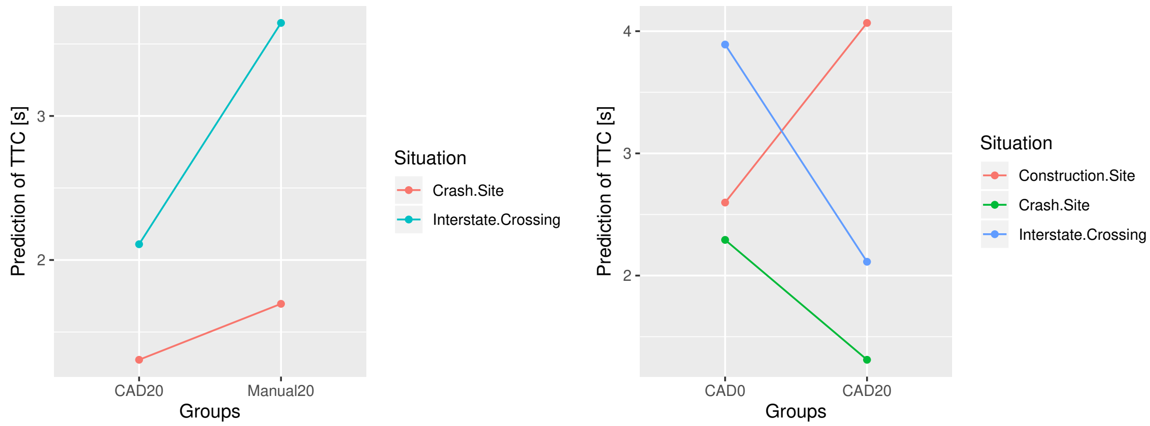


Figure A.5: Plot of the point estimates for the TTC revealing the significant effects for situation and the factor traffic density in the right plot.

Table A.7: Overview of test results on normal distribution and homogeneity of variance for the measures from the seat pressure mats.

Measure	Shapiro-Wilk	Levene	F_{max}
	COP – seat		
Groups	CAD0: $W = .885, p < .001$, CAD20: $W = .783, p < .001$	$F(1, 109) = .22, p = .64$	-
Duration	5A: $W = .867, p < .001$, 5B: $W = .746, p < .001$, 30: $W = .941, p = .05$	- (within)	-
	COP – backrest		
Groups	CAD0: $W = .932, p < .01$, CAD20: $W = .727, p < .001$	$F(1, 109) = .15, p = .70$	-
Duration	5A: $W = .977, p = .62$, 5B: $W = .901, p < .01$, 30: $W = .893, p < .01$	- (within)	-
	Contact area – seat		
Groups	CAD0: $W = .898, p < .001$, CAD20: $W = .718, p < .001$	$F(1, 103) = .49, p = .49$	-
Duration	5A: $W = .851, p < .001$, 5B: $W = .802, p < .001$, 30: $W = .785, p < .001$	- (within)	-
	Contact area – backrest		
Groups	CAD0: $W = .858, p < .001$, CAD20: $W = .816, p < .001$	$F(1, 104) = .94, p = .33$	-
Duration	5A: $W = .929, p = .02$, 5B: $W = .779, p < .001$, 30: $W = .875, p < .001$	- (within)	-

B Appendix for Chapter 7 "Modeling of take-over performance"

Table B.1: Results from checking all predictors for multicollinearity. The string "lastmin" was abbreviated to "lm". Significant associations are colored.

Predictor	Predictor	Ass.	Type	Obs. pairs
age	sex	0.15	ANOVA	299
years_driving	sex	0.16	ANOVA	299
km_peryear	sex	0.28	Cramer's V	299
subj_drivingstyle	sex	0.22	ANOVA	299
situation	sex	0.00	Cramer's V	299
ndrt	sex	0.11	Cramer's V	299
instruction	sex	0.00	Cramer's V	299
traffic	sex	0.03	ANOVA	299
hgd_lm	sex	0.02	ANOVA	290
hgd_lm2_perc	sex	0.06	ANOVA	276
peor_lm	sex	0.13	ANOVA	285
peor_10s	sex	0.07	ANOVA	299
peor_lm2_perc	sex	0.02	ANOVA	271
blinkdur_lm	sex	0.20	ANOVA	266
blinkdur_lm2_perc	sex	0.11	ANOVA	252
blinkfrequ_lm	sex	0.10	ANOVA	266
blinkfrequ_lm2_perc	sex	0.05	ANOVA	253
copseat_lm2_perc	sex	0.02	ANOVA	299
copback_lm2_perc	sex	0.02	ANOVA	299
years_driving	age	0.96	Corr.	299
km_peryear	age	0.46	ANOVA	299
subj_drivingstyle	age	-0.20	Corr.	299
situation	age	0.04	ANOVA	299
ndrt	age	0.08	ANOVA	299
instruction	age	0.04	ANOVA	299
traffic	age	-0.20	Corr.	299
hgd_lm	age	0.17	Corr.	290
hgd_lm2_perc	age	0.15	Corr.	276
peor_lm	age	0.02	Corr.	285
peor_10s	age	0.06	Corr.	299
peor_lm2_perc	age	-0.04	Corr.	271
blinkdur_lm	age	-0.03	Corr.	266
blinkdur_lm2_perc	age	-0.11	Corr.	252
blinkfrequ_lm	age	-0.21	Corr.	266
blinkfrequ_lm2_perc	age	-0.12	Corr.	253
copseat_lm2_perc	age	-0.03	Corr.	299
copback_lm2_perc	age	0.01	Corr.	299
km_peryear	years_driving	0.46	ANOVA	299
subj_drivingstyle	years_driving	-0.18	Corr.	299
situation	years_driving	0.03	ANOVA	299
ndrt	years_driving	0.07	ANOVA	299
instruction	years_driving	0.03	ANOVA	299
traffic	years_driving	-0.13	Corr.	299
hgd_lm	years_driving	0.18	Corr.	290
hgd_lm2_perc	years_driving	0.13	Corr.	276
peor_lm	years_driving	0.01	Corr.	285
peor_10s	years_driving	0.04	Corr.	299
peor_lm2_perc	years_driving	-0.02	Corr.	271
blinkdur_lm	years_driving	-0.03	Corr.	266
blinkdur_lm2_perc	years_driving	-0.08	Corr.	252
blinkfrequ_lm	years_driving	-0.22	Corr.	266
blinkfrequ_lm2_perc	years_driving	-0.13	Corr.	253
copseat_lm2_perc	years_driving	-0.02	Corr.	299
copback_lm2_perc	years_driving	-0.00	Corr.	299
subj_drivingstyle	km_peryear	0.29	ANOVA	299
situation	km_peryear	0.00	Cramer's V	299
ndrt	km_peryear	0.06	Cramer's V	299
instruction	km_peryear	0.03	Cramer's V	299
traffic	km_peryear	0.16	ANOVA	299
hgd_lm	km_peryear	0.07	ANOVA	290
hgd_lm2_perc	km_peryear	0.15	ANOVA	276
peor_lm	km_peryear	0.12	ANOVA	285
peor_10s	km_peryear	0.15	ANOVA	299
peor_lm2_perc	km_peryear	0.09	ANOVA	271
blinkdur_lm	km_peryear	0.17	ANOVA	266
blinkdur_lm2_perc	km_peryear	0.15	ANOVA	252
blinkfrequ_lm	km_peryear	0.28	ANOVA	266
blinkfrequ_lm2_perc	km_peryear	0.12	ANOVA	253
copseat_lm2_perc	km_peryear	0.08	ANOVA	299
copback_lm2_perc	km_peryear	0.14	ANOVA	299
situation	subj_drivingstyle	0.10	ANOVA	299
ndrt	subj_drivingstyle	0.19	ANOVA	299
instruction	subj_drivingstyle	0.10	ANOVA	299
traffic	subj_drivingstyle	0.02	Corr.	299
hgd_lm	subj_drivingstyle	-0.01	Corr.	290
hgd_lm2_perc	subj_drivingstyle	-0.07	Corr.	276
peor_lm	subj_drivingstyle	-0.13	Corr.	285
peor_10s	subj_drivingstyle	-0.09	Corr.	299
peor_lm2_perc	subj_drivingstyle	-0.04	Corr.	271
blinkdur_lm	subj_drivingstyle	0.04	Corr.	266
blinkdur_lm2_perc	subj_drivingstyle	0.04	Corr.	252
blinkfrequ_lm	subj_drivingstyle	-0.06	Corr.	266
blinkfrequ_lm2_perc	subj_drivingstyle	-0.05	Corr.	253
copseat_lm2_perc	subj_drivingstyle	-0.12	Corr.	299
copback_lm2_perc	subj_drivingstyle	-0.08	Corr.	299
ndrt	situation	0.28	Cramer's V	299
instruction	situation	0.24	Cramer's V	299
traffic	situation	0.31	ANOVA	299
hgd_lm	situation	0.19	ANOVA	290
hgd_lm2_perc	situation	0.16	ANOVA	276
peor_lm	situation	0.15	ANOVA	285
peor_10s	situation	0.21	ANOVA	299
peor_lm2_perc	situation	0.12	ANOVA	271
blinkdur_lm	situation	0.14	ANOVA	266
blinkdur_lm2_perc	situation	0.09	ANOVA	252
blinkfrequ_lm	situation	0.04	ANOVA	266
blinkfrequ_lm2_perc	situation	0.05	ANOVA	253
copseat_lm2_perc	situation	0.18	ANOVA	299
copback_lm2_perc	situation	0.08	ANOVA	299
instruction	ndrt	0.32	Cramer's V	299
traffic	ndrt	0.54	ANOVA	299
hgd_lm	ndrt	0.34	ANOVA	290
hgd_lm2_perc	ndrt	0.23	ANOVA	276
peor_lm	ndrt	0.82	ANOVA	285
peor_10s	ndrt	0.79	ANOVA	299
peor_lm2_perc	ndrt	0.63	ANOVA	271
blinkdur_lm	ndrt	0.28	ANOVA	266
blinkdur_lm2_perc	ndrt	0.20	ANOVA	252
blinkfrequ_lm	ndrt	0.57	ANOVA	266
blinkfrequ_lm2_perc	ndrt	0.56	ANOVA	253
copseat_lm2_perc	ndrt	0.25	ANOVA	299
copback_lm2_perc	ndrt	0.13	ANOVA	299
traffic	instruction	0.33	ANOVA	299
hgd_lm	instruction	0.11	ANOVA	290
hgd_lm2_perc	instruction	0.01	ANOVA	276
peor_lm	instruction	0.17	ANOVA	285
peor_10s	instruction	0.22	ANOVA	299
peor_lm2_perc	instruction	0.03	ANOVA	271
blinkdur_lm	instruction	0.09	ANOVA	266
blinkdur_lm2_perc	instruction	0.01	ANOVA	252
blinkfrequ_lm	instruction	0.07	ANOVA	266
blinkfrequ_lm2_perc	instruction	0.00	ANOVA	253
copseat_lm2_perc	instruction	0.06	ANOVA	299
copback_lm2_perc	instruction	0.12	ANOVA	299
hgd_lm	traffic	-0.08	Corr.	290
hgd_lm2_perc	traffic	-0.03	Corr.	276
peor_lm	traffic	-0.17	Corr.	285
peor_10s	traffic	-0.21	Corr.	299
peor_lm2_perc	traffic	-0.17	Corr.	271
blinkdur_lm	traffic	-0.08	Corr.	266
blinkdur_lm2_perc	traffic	0.06	Corr.	252
blinkfrequ_lm	traffic	0.02	Corr.	266
blinkfrequ_lm2_perc	traffic	-0.03	Corr.	253
copseat_lm2_perc	traffic	-0.08	Corr.	299
copback_lm2_perc	traffic	-0.09	Corr.	299
hgd_lm2_perc	hgd_lm	0.68	Corr.	276
peor_lm	hgd_lm	-0.36	Corr.	285
peor_10s	hgd_lm	-0.25	Corr.	290
peor_lm2_perc	hgd_lm	-0.18	Corr.	271
blinkdur_lm	hgd_lm	-0.18	Corr.	266
blinkdur_lm2_perc	hgd_lm	-0.14	Corr.	252
blinkfrequ_lm	hgd_lm	-0.06	Corr.	266
blinkfrequ_lm2_perc	hgd_lm	-0.11	Corr.	252
copseat_lm2_perc	hgd_lm	0.05	Corr.	290
copback_lm2_perc	hgd_lm	-0.05	Corr.	290

peor_lm	hgd_lm2_perc	-0.37	Corr.	271
peor_10s	hgd_lm2_perc	-0.23	Corr.	276
peor_lm2_perc	hgd_lm2_perc	-0.39	Corr.	271
blinkdur_lm	hgd_lm2_perc	-0.16	Corr.	254
blinkdur_lm2_perc	hgd_lm2_perc	-0.17	Corr.	252
blinkfrequ_lm	hgd_lm2_perc	-0.07	Corr.	254
blinkfrequ_lm2_perc	hgd_lm2_perc	-0.16	Corr.	252
copseat_lm2_perc	hgd_lm2_perc	0.07	Corr.	276
copback_lm2_perc	hgd_lm2_perc	0.01	Corr.	276
peor_10s	peor_lm	0.83	Corr.	285
peor_lm2_perc	peor_lm	0.75	Corr.	271
blinkdur_lm	peor_lm	0.32	Corr.	266
blinkdur_lm2_perc	peor_lm	0.18	Corr.	252
blinkfrequ_lm	peor_lm	0.35	Corr.	266
blinkfrequ_lm2_perc	peor_lm	0.40	Corr.	252
copseat_lm2_perc	peor_lm	0.07	Corr.	285
copback_lm2_perc	peor_lm	0.02	Corr.	285
peor_lm2_perc	peor_10s	0.67	Corr.	271
blinkdur_lm	peor_10s	0.30	Corr.	266
blinkdur_lm2_perc	peor_10s	0.15	Corr.	252
blinkfrequ_lm	peor_10s	0.28	Corr.	266
blinkfrequ_lm2_perc	peor_10s	0.37	Corr.	253
copseat_lm2_perc	peor_10s	0.11	Corr.	299
copback_lm2_perc	peor_10s	0.05	Corr.	299
blinkdur_lm	peor_lm2_perc	0.19	Corr.	254
blinkdur_lm2_perc	peor_lm2_perc	0.11	Corr.	252
blinkfrequ_lm	peor_lm2_perc	0.41	Corr.	254
blinkfrequ_lm2_perc	peor_lm2_perc	0.54	Corr.	252
copseat_lm2_perc	peor_lm2_perc	0.08	Corr.	271
copback_lm2_perc	peor_lm2_perc	0.03	Corr.	271
blinkdur_lm2_perc	blinkdur_lm	0.69	Corr.	252
blinkfrequ_lm	blinkdur_lm	0.15	Corr.	266
blinkfrequ_lm2_perc	blinkdur_lm	0.19	Corr.	252
copseat_lm2_perc	blinkdur_lm	-0.04	Corr.	266
copback_lm2_perc	blinkdur_lm	0.10	Corr.	266
blinkfrequ_lm	blinkdur_lm2_perc	0.18	Corr.	252
blinkfrequ_lm2_perc	blinkdur_lm2_perc	0.20	Corr.	252
copseat_lm2_perc	blinkdur_lm2_perc	-0.10	Corr.	252
copback_lm2_perc	blinkdur_lm2_perc	0.04	Corr.	252
blinkfrequ_lm2_perc	blinkfrequ_lm	0.69	Corr.	252
copseat_lm2_perc	blinkfrequ_lm	0.03	Corr.	266
copback_lm2_perc	blinkfrequ_lm	-0.02	Corr.	266
copseat_lm2_perc	blinkfrequ_lm2_perc	0.14	Corr.	253
copback_lm2_perc	blinkfrequ_lm2_perc	0.02	Corr.	253
copback_lm2_perc	copseat_lm2_perc	-0.08	Corr.	299

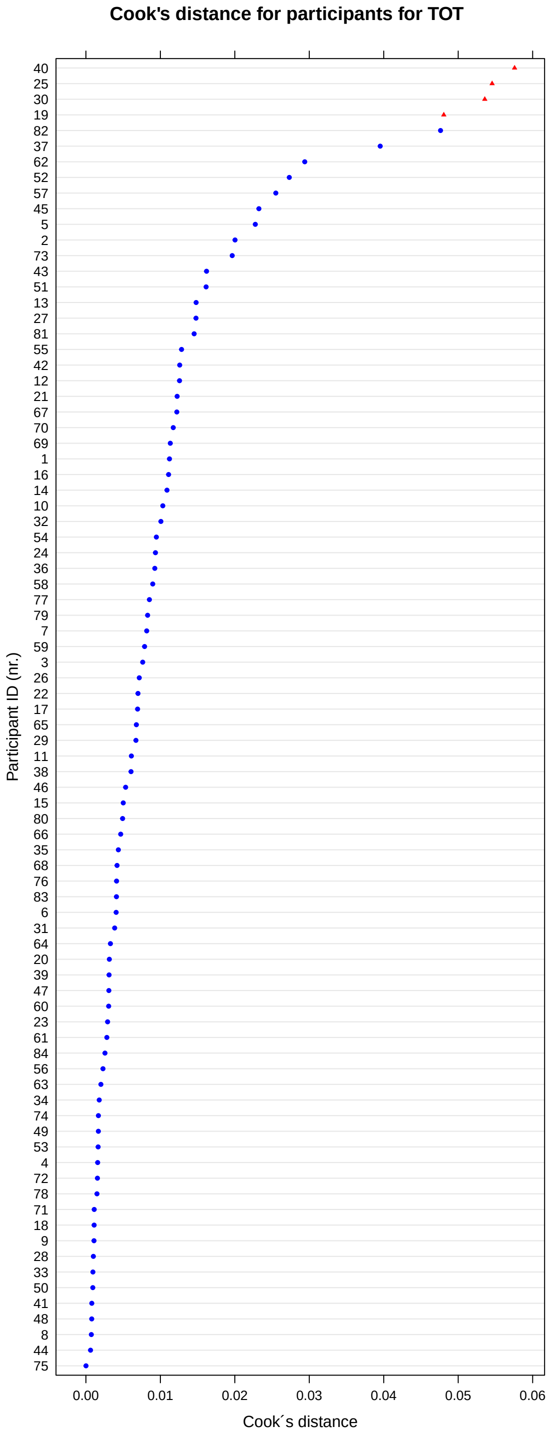


Figure B.1: Plot of the Cook's distances for the final model of the TOT. Participants 19, 25, 30 and 40 were identified as influential points.

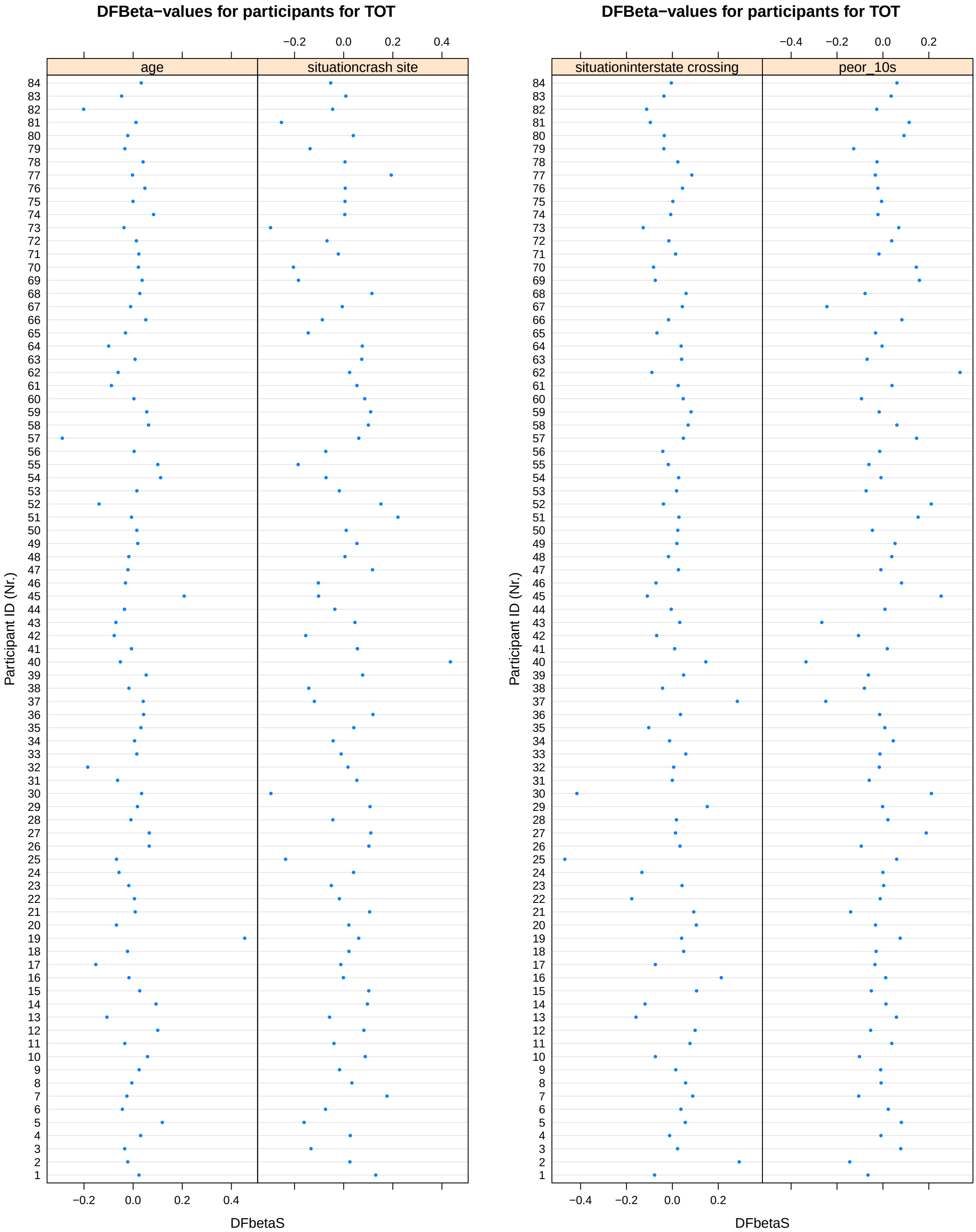
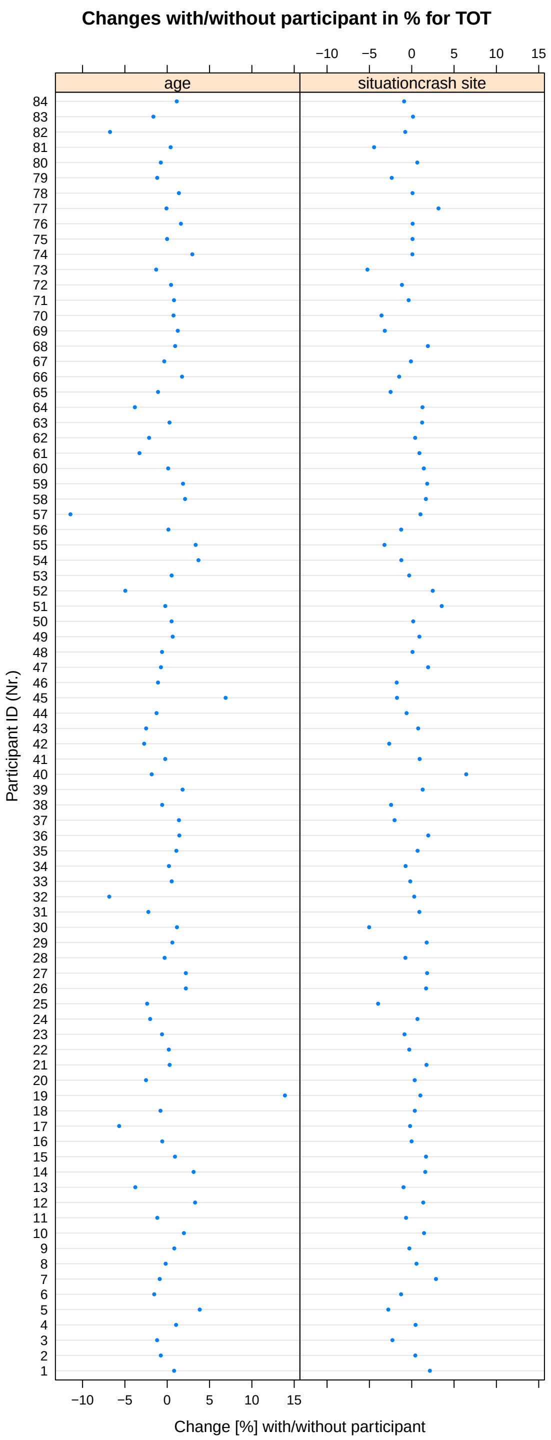


Figure B.2: Plot of the DFBetas for the TOT.



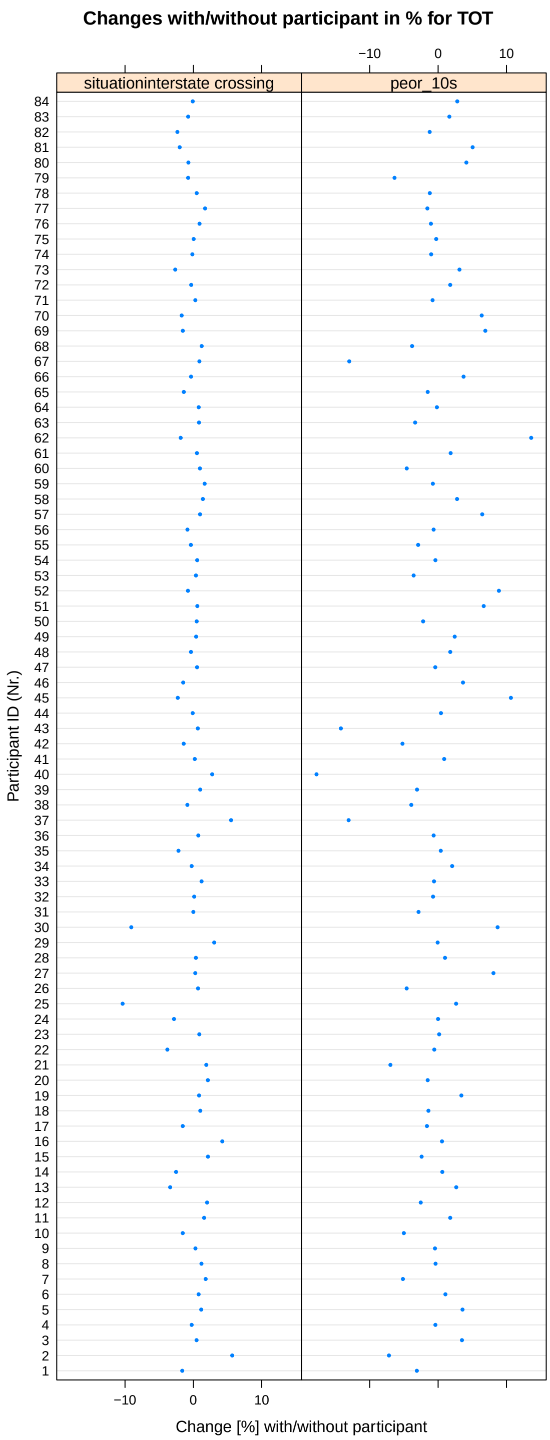


Figure B.3: Plot of the "percentage of change [...] as the absolute difference between the parameter estimate both including and excluding the higher-level unit" (Nieuwenhuis et al., 2012) for the TOT. The higher level unit in this work is the individual participant.

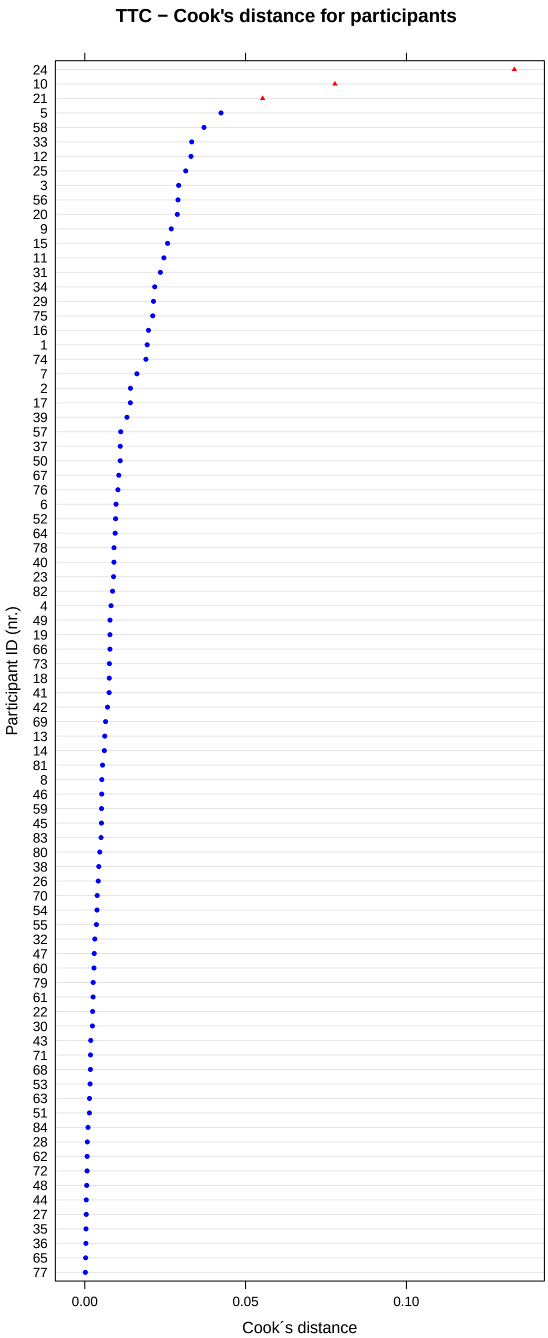


Figure B.4: Plot of the Cook's distances for the final model of the TTC. Participants 10, 21 and 24 were identified as influential points.

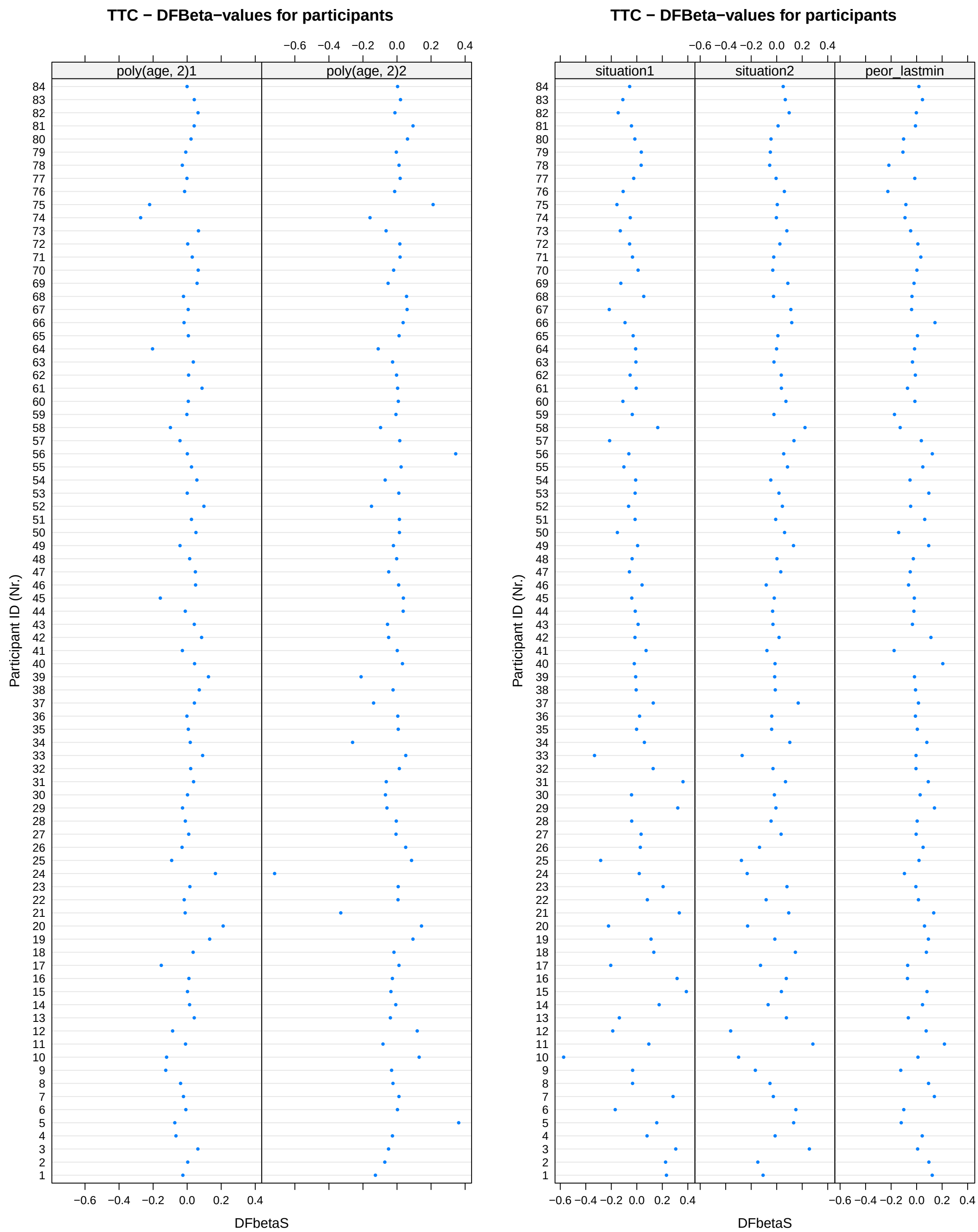


Figure B.5: Plot of the DFBetas for the TTC.

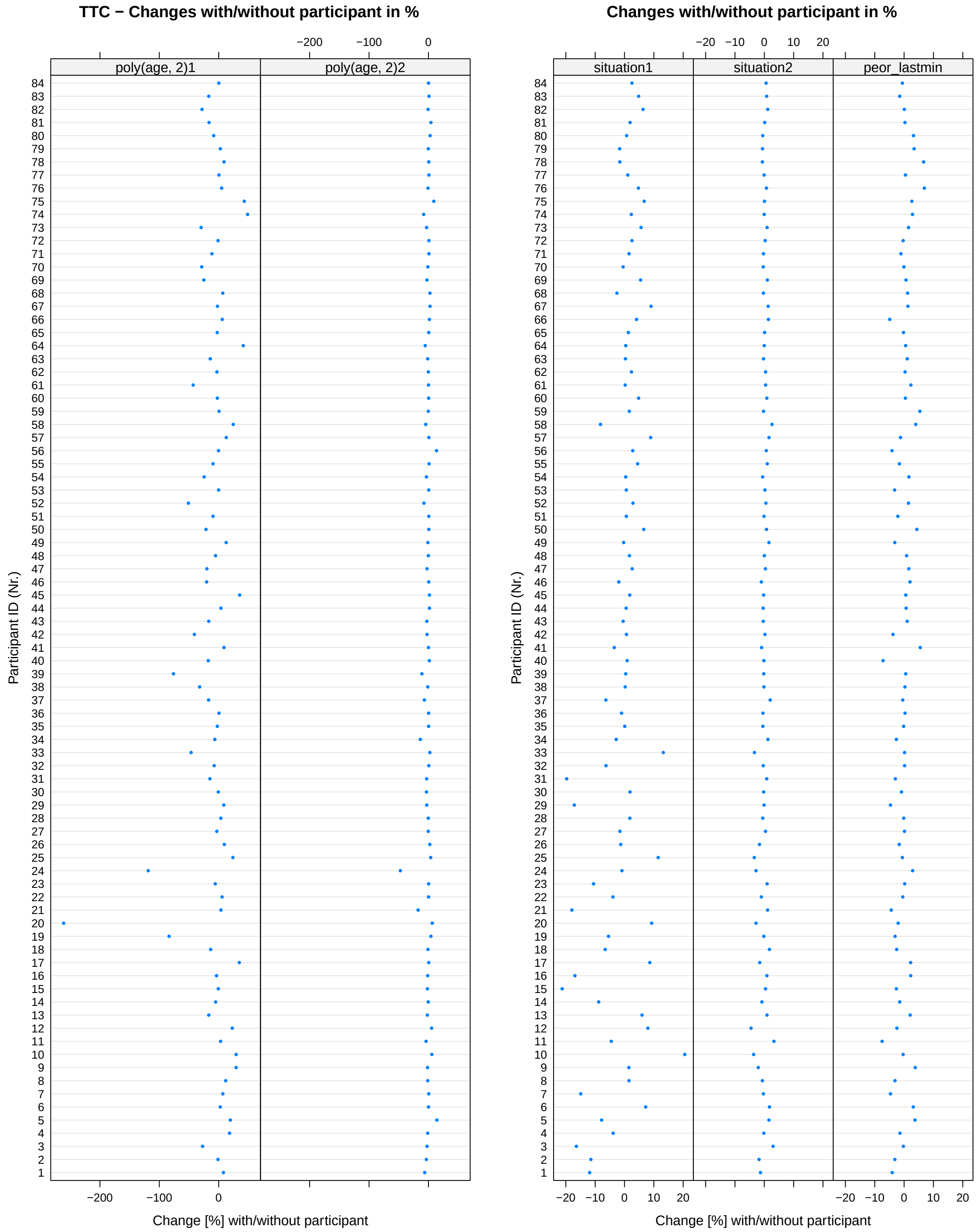


Figure B.6: Plot of "the percentage of change [...] as the absolute difference between the parameter estimate both including and excluding the higher-level unit" (Nieuwenhuis et al., 2012) for the TTC. The higher level unit in this work is the individual participant.

C Appendix for Chapter 9 "Design and evaluation of an optimized human-machine interface for the take-over"

Table C.1: Overview of test results on normal distribution and homogeneity of variance for the objective take-over performance and the eye-tracking data.

Dependent Variable	Shapiro-Wilk	Levene	F_{max}
Take-over time (TOT)	No HUD: $W = .936, p = .03$, With HUD: $W = .934, p = .05$, Construction Site: $W = .948, p = .01$, Tight Curve: $W = .932, p = .03$	$F(1,69) = .06, p = .81$	-
Lateral acceleration	No HUD: $W = .947, p = .06$, With HUD: $W = .854, p < .01$, Construction Site: $W = .918, p < .01$, Tight Curve: $W = .908, p < .01$	$F(1,78) = 2.10, p = .15$	-
Standard deviation of lateral position (SDLP)	No HUD: $W = .595, p < .001$, With HUD: $W = .479, p < .001$, Construction Site: $W = .514, p < .001$, Tight Curve: $W = .566, p < .001$	$F(1,72) = 2.55, p = .11$	-
Percentage eyes on road (PEOR)	No HUD: $W = .967, p = .42$, With HUD: $W = .957, p = .19$, Construction Site: $W = .973, p = .56$, Tight Curve: $W = .956, p = .19$	$F(1,64) = .47, p = .49$	-
Percentage eyes on instrument cluster (PEOIC)	No HUD: $W = .926, p = .03$, With HUD: $W = .718, p < .01$, Construction Site: $W = .890, p < .01$, Tight Curve: $W = .832, p < .01$	$F(1,64) = 2.74, p = .10$	-

Table C.2: Overview of test results on normal distribution and homogeneity of variance for the subjective ratings of the take-over situations.

Dependent Variable	Shapiro-Wilk	Levene	F_{max}
Criticality	No HUD: $W = .819, p < .001$, With HUD: $W = .846, p < .001$, Construction Site: $W = .890, p < .001$, Tight Curve: $W = .745, p < .001$	$F(1,78) = 4.60, p = .04$	2.5
Complexity	No HUD: $W = .840, p < .001$, With HUD: $W = .880, p < .001$, Construction Site: $W = .880, p < .001$, Tight Curve: $W = .808, p < .001$	$F(1,78) = 3.50, p = .07$	-
Time Budget	No HUD: $W = .816, p < .001$, With HUD: $W = .767, p < .001$, Construction Site: $W = .858, p < .001$, Tight Curve: $W = .691, p < .001$	$F(1,78) = 1.16, p = .28$	-
Obviousness	No HUD: $W = .840, p < .001$, With HUD: $W = .781, p < .001$, Construction Site: $W = .690, p < .001$, Tight Curve: $W = .872, p < .001$	$F(1,78) = 12.62, p < .001$	1.9

Table C.3: Overview of test results on normal distribution and homogeneity of variance for the subjective ratings of the HMI after each situation.

Dependent Variable	Shapiro-Wilk	Levene	F_{max}
Usefulness	No HUD: $W = .875, p < .001$, With HUD: $W = .855, p < .001$, Construction Site: $W = .895, p < .01$, Tight Curve: $W = .879, p < .001$	$F(1,78) = 5.65, p = .02$	2.2
Satisfaction	No HUD: $W = .905, p < .01$, With HUD: $W = .732, p < .001$, Construction Site: $W = .874, p < .001$, Tight Curve: $W = .898, p < .01$	$F(1,78) = 3.90, p = .05$	-

Table C.4: Overview of test results on normal distribution and homogeneity of variance for the subjective ratings of the HMI in the final questionnaire.

Dependent Variable	Shapiro-Wilk	Levene	F_{max}
Safety	No HUD: $W = .930, p = .16$, With HUD: $W = .933, p = .17$	$F(1, 38) = 5.34, p = .03$	2.3
Usability	No HUD: $W = .933, p = .17$, With HUD: $W = .895, p = .03$	$F(1, 38) = 5.87, p = .02$	2.7
Intention to use	No HUD: $W = .909, p = .06$, With HUD: $W = .916, p = .08$	$F(1, 38) = 4.49, p = .04$	1.9