Enhancing User Experience and Safety
in the Context of Automated Driving
Through Uncertainty Communication

Loughborough University

DOCTORAL THESIS

Enhancing User Experience and Safety in the Context of Automated Driving Through Uncertainty Communication

Interface Design for Vehicles Equipped with Automated Driving Systems

Authored by

Alexander Kunze

Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy (PhD) of Loughborough University, July 2019.

Author
Alexander Kunze
Loughborough University
School of Design and Creative Arts
Loughborough, United Kingdom
European Union

This thesis was typeset using $\LaTeX\,2_\varepsilon$.
The R environment was used for data handling, analysis, and visualisation.

Automation does not replace human performance,
it changes human performance.

Raja Parasuraman
University Professor

Acknowledgements

Where there is order now there was chaos in the beginning. Navigating through the seemingly endless body of existing research, it was soon blindingly obvious what Bernard de Chartres meant by *standing on the shoulders of giants*. Through generous measures of perseverance and, frankly, time, initially concealed interrelations between individual research items became apparent and shortly after, gaps in knowledge revealed themselves.

Among all virtues tried and developed in this endeavour, the arguably most significant one shall not remain unaddressed: gratitude. As such, the constant support of my supervisors has not gone unnoticed and shall be rewarded with my utmost appreciation. Thank you, Steve, Ashleigh, and Russell, for providing words of criticism where appropriate and gestures of encouragement when needed. In this context, I am also indebted to Professor Andrew Morris, who served as independent assessor throughout this research project and provided further inspiration from a different perspective. Moreover, the lively discussion during the viva led to several improvements that helped to future-proof this thesis – thank you, Professor Diane Gyi and Professor Gary Burnett.

Words of gratitude shall also be extended to my fellow doctoral students. Daily if not constant discussions sparked new ideas and provided unfiltered feedback that shaped the work presented in this thesis.

While this project truly was intellectually stimulating, keeping motivated during the whole doctoral journey was probably the greatest challenge. In this regard, particularly the warden team of the iconic Towers accommodation, Ruth, Matt, Leah, Karol, and Dimitris, ensured an enjoyable work-life balance.

Most important of all, the support and affection shown by my family, partner, and closest friends provided never-ceasing tailwind that pushed me forward.

Preface

Uncertainties are not a novel concept for humans. To various degrees, uncertainties pervade every aspect of human life. At the most elemental level, uncertainties delineate the unknown, the unpredictable, and the indefinite. Without exception, future events are associated with uncertainties that render the exact determination of forthcoming states impossible. Similarly, most of the past cannot be recovered with complete confidence and likewise in the present, required information remains partially hidden or unattainable. Consciously or subliminally, humans perpetually account for uncertainties. Our decisions and actions are, in varying degrees, affected by the uncertainties of the underlying information. Stockbrokers, for instance, incorporate uncertainties about the development of equity prices into purchasing decisions, parents consider the uncertainty of the weather forecast when dressing their children in the morning, and drivers reduce their speed in uncertain environments, e.g. in cases of heavy fog or snowfall.

During interactions, the uncertainties pertaining to the characteristics and abilities of the other individual are incorporated into our decision to *trust* or *distrust* someone. This becomes particularly apparent when delegating tasks to other agents. Imagine you find yourself on the way to a distant destination, seated on the passenger seat of a manually driven vehicle. A long-time friend – whom you have come to appreciate as a very reliable and safe driver – controls the vehicle. Unless you are a hypercritical passenger, you will, in all likelihood, refrain from monitoring her every movement. One might say *you trust in her abilities to safely perform the driving task* – a trust that has been built based on a long-lasting relationship. Contrasting this, imagine the driver next to you as a juvenile stranger who is visibly anxious to perform the task at hand. Whereas the previous scenario implied a relaxed journey, the current situation likely prompts you to closely monitor the driving performance of your chauffeur – or you might even request to take over control of the vehicle yourself. While you were fairly certain that you would safely arrive at the destination when travelling with your friend, the inexperienced, anonymous driver of the second scenario would leave you uncertain about the outcome of this journey.

Now consider a third scenario that, once again, leaves you as a passive passenger. Unlike before, you find yourself on the driver seat. You are, however, not performing the driving task – the vehicle is controlled by an automated system. While the previous scenarios allowed you to gauge your trust in the other agents, the automated system is a completely opaque black box. Is it more akin to the former, very reliable and trustworthy driver or does it perform more like the latter, highly inexperienced driver? For many years to come, the answer will be: *it depends.*

Similar to humans, automated systems are affected by uncertainties. Erroneous sensor inputs or environmental factors such as fog might, for instance, contribute to these uncertainties. In situations characterised by low uncertainties, the automated system will perform in a superior manner compared with any human driver. Equipped with numerous sensors that offer a 360 degrees field of view around the vehicle, automated systems achieve exceptionally fast and accurate responses to changes in the environment. In situations of high uncertainty, however, the system might not be able to correctly interpret changes in the environment and might fail to determine a suitable response. In these situations, the automated system relies on the help of the human user – who was completely disengaged from the driving task a moment ago and is likely taken by surprise.

How should users know when the system can cope with the situation at hand and when it cannot? Can a driver be expected to memorise the operational limits of an automated system – limits that not even the developers themselves might be aware of? The answer to this question must be *no*.

Imperfect automated driving systems that rely on the human operator to take over the driving task in cases of system failures must be designed to facilitate system understanding and awareness. Even the most novice drivers must be able to accurately and effortlessly determine the likelihood with which situations fall outside of safe operation boundaries.

The research project reported in this thesis explores the dynamic communication of system uncertainties as a means for achieving this. Revealing the uncertainties of the automated driving system reduces the uncertainties of the human operators regarding the current reliability of the automated system and facilitates a more accurate judgement of the system's current trustworthiness. By rendering the previously black box more transparent, users are better able to anticipate system failures and can adapt their monitoring behaviour accordingly. The following pages will visit this approach for overcoming the inherent human factors challenges that the automation of the driving task entails in more detail – challenges which will persist until the automobile truly becomes an *auto*mobile.

Abstract

Operators of highly automated driving systems may exhibit behaviour characteristic of overtrust issues due to an insufficient awareness of automation fallibility. Consequently, situation awareness in critical situations is reduced and safe driving performance following emergency takeovers is impeded. Previous research has indicated that conveying system uncertainties may alleviate these issues. However, existing approaches require drivers to attend the uncertainty information with focal attention, likely resulting in missed changes when engaged in non-driving-related tasks.

This research project expands on existing work regarding uncertainty communication in the context of automated driving. Specifically, it aims to investigate the implications of conveying uncertainties under consideration of non-driving-related tasks and, based on the outcomes, develop and evaluate an uncertainty display that enhances both user experience and driving safety. In a first step, the impact of visually conveying uncertainties was investigated under consideration of workload, trust, monitoring behaviour, non-driving-related tasks, takeover performance, and situation awareness. For this, an anthropomorphic visual uncertainty display located in the instrument cluster was developed. While the hypothesised benefits for trust calibration and situation awareness were confirmed, the results indicate that visually conveying uncertainties leads to an increased perceived effort due to a higher frequency of monitoring glances.

Building on these findings, peripheral awareness displays were explored as a means for conveying uncertainties without the need for focused attention to reduce monitoring glances. As a prerequisite for developing such a display, a systematic literature review was conducted to identify evaluation methods and criteria, which were then coerced into a comprehensive framework. Grounded in this framework, a peripheral awareness display for uncertainty communication was developed and subsequently compared with the initially proposed visual anthropomorphic uncertainty display in a driving simulator study. Eye tracking and subjective workload data indicate that the peripheral awareness display reduces the monitoring effort relative to the visual display, while driving performance and trust data highlight that the benefits of uncertainty communication are maintained.

Further, this research project addresses the implications of increasing the functional detail of uncertainty information. Results of a driving simulator study indicate that particularly workload should be considered when increasing the functional detail of uncertainty information. Expanding upon this approach, an augmented reality display concept was developed and a set of visual variables was explored in a forced choice sorting task to assess their ordinal characteristics. Particularly changes in colour hue and animation-based variables received high preference ratings and were ordered consistently from low to high uncertainty.

This research project has contributed a series of novel insights and ideas to the field of human factors in automated driving. It confirmed that conveying uncertainties improves trust calibration and situation awareness, but highlighted that using a visual display lessens the positive effects. Addressing this shortcoming, a peripheral awareness display was designed applying a dedicated evaluation framework. Compared with the previously employed visual display, it decreased monitoring glances and, consequentially, perceived effort. Further, an augmented reality-based uncertainty display concept was developed to minimise the workload increments associated with increases in the functional detail of uncertainty information.

Keywords Automated driving; uncertainty; trust in automation; situation awareness; interaction design; human-machine interface.

Publications

The contents of this thesis are published in the following, chronologically ordered academic journal (VI, VII) and conference papers (I – IV, VIII). Further, the outcomes of this research project have contributed to two workshops on trust in automation (V, IX).

[I] **Kunze**, Alexander et al. (2017). *'Enhancing Driving Safety and User Experience Through Unobtrusive and Function-Specific Feedback'*. In: *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. DOI: 10.1145/3131726.3131762.

[II] **Kunze**, Alexander et al. (2018). *'Augmented Reality Displays for Communicating Uncertainty Information in Automated Driving'*. In: *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. DOI: 10.1145/3239060.3239074.

[III] **Kunze**, Alexander et al. (2018). *'Evaluation of Variables for the Communication of Uncertainties Using Peripheral Awareness Displays'*. In: *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. DOI: 10.1145/3239092.3265958.

[IV] **Kunze**, Alexander et al. (2018). *'Preliminary Evaluation of Variables for Communicating Uncertainties Using a Haptic Seat'*. In: *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. DOI: 10.1145/3239092.3265959.

[V] Wintersberger, Philipp, Alexander G. Mirnig, Brittany E. Noah, Alexander **Kunze**, Johannes Kraus, Shailie Thakkar, Roderick McCall, Bruce N. Walker (2018). *'Second Workshop on Trust in the Age of Automated Driving'*. In: *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. DOI: 10.1145/3239092.3265959.

[VI] **Kunze**, Alexander et al. (2019). *'Automation Transparency: Implications of Uncertainty Communication for Human-Automation Interaction and Interfaces'*. In: *Ergonomics* 62.3. DOI: 10.1080/00140139.2018.1547842.

[VII] **Kunze**, Alexander et al. (2019). *'Function-Specific Uncertainty Communication in Automated Driving'*. In: *Special Issue of the International Journal of Mobile Human Computer Interaction (IJMHCI): Recent Advances in Automotive User Interfaces and Interactive Vehicular Applications Research* 11.2. DOI: 10.4018/IJMHCI.2019040105.

[VIII] **Kunze**, Alexander et al. (2019). *'Conveying Uncertainties Using Peripheral Awareness Displays in the Context of Automated Driving'*. In: *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. DOI: 10.1145/3342197.3344537.

[IX] Holthausen, Brittany, Philipp Wintersberger, Zoe M. Becerra, Alexander G. Mirnig, Alexander **Kunze**, Bruce N. Walker (2019). *'Third Workshop on Trust in Automation: How Does Trust Influence Interaction'*. In: *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. DOI: 10.1145/3349263.3350761.

Contents

Acronyms

ABS . . . . . . anti-lock braking system

ACC . . . . . . adaptive cruise control

ADS . . . . . . automated driving system

ADS-DV . . . . ADS-dedicated vehicle

AI . . . . artificial intelligence

ANN . . . . . . artificial neural networks

ANOVA . . . . analysis of variance

AOI . . . . area of interest

AR . . . . augmented reality

BASt . . . . . . Bundesanstalt für Straßenwesen

caHUD . . . . . contact analogue head-up display

CbW . . . . . . Conduct-by-Wire

CC . . . . cruise control

CID . . . . combat identification

CIE . . . . International Commission on Illumination

CUE . . . . . . Components of User Experience

CUEPD . . . . Context of Use Evaluation of Peripheral Displays

DC . . . . direct current

DDT . . . . . . dynamic driving task

DL . . . . deep learning

DNN . . . . . . deep neural network

DRT . . . . . . driving-related task

DSI . . . . Driver Skill Inventory

DVAS . . . . . discrete visual analogue scale

ECG . . . . . . electrocardiogram

EDA . . . . . . electrodermal activity

EMG . . . . . . eletromyography

ERTRAC . . . European Road Transport Research Advisory Council

ESC . . . . electronic stability control

FA . . . . false alarm

FAA . . . . . . Federal Aviation Administration

FAE . . . . false alarm effect

FOV . . . . . . field of view

FRD . . . . . . field relevant for driving

GIS . . . . geographic information system

GIScience . . . geographic information science

GM . . . . General Motors

GPS . . . . . . Global Positioning System

GPU . . . . . . graphics processing unit

H-mode . . . . Horse-mode

HASO . . . . . Human-Autonomy System Oversight

HCI . . . . human-computer interaction

HF . . . . human factors

HGV . . . . . . heavy goods vehicle

HMD . . . . . . head-mounted display

HMI . . . . . . human-machine interface

HRV . . . . . . heart rate variability

HSL . . . . hue, saturation, lightness

HUD . . . . . . head-up display

IAE . . . . information access effort

IMU . . . . . . inertial measurement unit

IR . . . . infrared

IRC . . . . Interruption, Reaction and Comprehension

ISO . . . . International Organization for Standardization

LMM . . . . . . linear mixed model

ML . . . . machine learning

MOSFET . . . metal-oxide semiconductor field-effect transistor

MTTC . . . . . minimum time to collision

NASA-TLX . . NASA Task Load Index

NDRT . . . . . non-driving-related task

NHTSA . . . . National Highway Traffic Safety Administration

NTSB . . . . . National Transportation Safety Board

ODD . . . . . . operational design domain

OECD . . . . . Organisation for Economic Co-operation and Development

OEDR . . . . . object and event detection and response

OEM . . . . . . original equipment manufacturer

OICA . . . . . . International Organisation of Motor Vehicle Manufacturers

OOTL . . . . . out-of-the-loop

PADE . . . . . Peripheral Awareness Display Evaluation

PCB . . . . . . printed circuit board

PDT . . . . . . percent dwell time

PMD . . . . . . Photonic Mixing Device

PWM . . . . . pulse width modulation

QUIM . . . . . Quality in Use Integrated Measurement

RGB . . . . . . red, green, blue

RMSSD . . . . root mean square of successive differences

RQ . . . . research question

RT . . . . response time

SA . . . . situation awareness

SAE . . . . Society of Automotive Engineers

SAGAT . . . . Situation Awareness Global Assessment Technique

SAM . . . . . . Self-Assessment Manikin

SART . . . . . Situation Awareness Rating Technique

SEEV . . . . . . salience, effort, expectancy, and value

SHERPA . . . Systematic Human Error Reduction and Prediction Approach

STSS . . . . . . short term sensory store

TH . . . . time headway

TOR . . . . . . take-over request

TTC . . . . . . time to collision

TTT . . . . . . time to takeover

UD . . . . uncertainty display

UI . . . . user interface

UK . . . . United Kingdom

UNECE . . . . United Nations Economic Commission for Europe

US . . . . United States

UV . . . . ultraviolet

UWP . . . . . . Universal Windows Platform

UX . . . . user experience

V2I . . . . vehicle-to-infrastructure

V2V . . . . . . vehicle-to-vehicle

WSD . . . . . . windshield display

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Motivation and Problem

Vehicles equipped with increasingly automated driving systems promise advances in safety and efficiency while simultaneously affording users the flexibility to engage in non-driving-related tasks (NDRTs). It is therefore not surprising that driving automation is among the key trends in technology (Gartner 2016) and will profoundly alter the way we live (Economist 2018). In 2016, 1.35 million road traffic deaths were reported globally. Among people aged 5 to 29 years, road traffic injuries are the leading cause of death (World Health Organization 2018). Considering that human errors account for more than 90 % of all roadway collisions (NHTSA 2008, pp. 23–25) and lead to annual costs of approximately $ 190 billion to the US economy alone (Bertoncello and Wee 2015), automating the driving task as much as possible appears to be the logical approach to the road safety problem.

An increasing degree of automation, however, does not necessarily correlate with a reduction in the importance of the human user. In fact, 'the more advanced a control system is, so the more crucial may be the contribution of the human operator (Bainbridge 1983, p. 775)'. This is known as the *irony of automation* and emphasises on the human factors (HF) challenges that have to be met before the anticipated safety benefits can be achieved (Trimble et al. 2014, p. 1). Rather than simply replacing human tasks, automation alters them – often to the disadvantage of human operators. Increasing degrees of automation leave users to assume the role of a passive monitor, a task humans are particularly inept for (Norman 2015, pp. 81–84). Operator intervention is reserved for complex situations that prompt the automated system to disengage or fail, i. e. circumstances in which users would benefit most from a functioning automated aid.

To avoid these HF challenges, some companies, e. g. Google, employ a revolutionary approach that promotes an outright shift from manual to fully automated driving. In contrast to the competing evolutional approach, however, it will be challenging to achieve sufficient levels of robustness and reliability (Endsley 2017; Woods 2016).

For the foreseeable future, users will therefore likely continue to be involved in the driving task, particularly in critical situations that exceed the capabilities of the automated system (Endsley 2017; Gomes 2016; European Road Transport Research Advisory Council 2015).

In such scenarios, the system issues a take-over request (TOR) to signal that automation disengagement is imminent. TORs require users to immediately focus on the driving scene and rapidly comprehend the situation after possibly being completely *out of the loop*, i. e. not involved in the driving task.

Enabling users to safely perform the driving task following takeovers by keeping them aware of system limitations is one of the key HF challenges researchers face in regard to vehicle automation (Kyriakidis et al. 2017). To prepare users for takeovers, previous research has suggested to constantly communicate system uncertainties (Beller, Heesen and Vollrath 2013; Helldin et al. 2013). Revealing the otherwise covert uncertainties of automated systems allows users to gauge the likelihood of TORs and is consequently hypothesised to support the acquisition and maintenance of situation awareness (SA) while the automation is engaged. Awareness of the current reliability of automated systems helps users with calibrating their trust in automation and results in more appropriate attention allocation strategies that facilitate the build-up of SA prior to critical situations (Hoff and Bashir 2015; Hergeth et al. 2016; Beller, Heesen and Vollrath 2013).

Besides benefits to TOR performance, knowledge of system uncertainties is expected to more generally improve human-automation cooperation as it establishes the automated system as a trustworthy agent and enables users to gauge who is currently more competent to perform the driving task. For instance, users may actively choose to drive before emergency takeovers are necessary. This would apply to all automated driving systems that allow the driver to perform the driving task. Further, the provision and tracking of system uncertainties can address questions of liability. In current vehicles, the legal responsibility rests with the driver at all times (Mirnig et al. 2019). Future vehicles will allow the user to disengage from the driving task in predefined operational domains for which the vehicle has been developed, for instance motorways. In these scenarios, liability will – in the future – likely rest with the entity behind the automated system (Law Commission of England and Wales and Scottish Law Commission 2018). However, dynamically changing environments, potential exclusions such as construction sites, or frequent software updates will make it difficult to convey the operational limits to the user. As such, users might not be aware of their liability for legal offences resulting from the driving task. Communicating system uncertainties renders the automation limitations, and the legal responsibility, visible to users and promotes the gradual development of an appropriate mental model without the necessity of extensive manuals.

The existing proposals for uncertainty communication mainly rely on visual information presented in the instrument cluster and thus require users to regularly shift their attention from NDRTs to the uncertainty display, even in safe driving situations (Beller, Heesen and Vollrath 2013; Helldin et al. 2013). This poses additional monitoring demands that increase the likelihood of missed events and entail regular user disruptions, thus increasing workload and impairing both safety and user experience. Research question (RQ) 1 was formulated to address this in more detail (see Section 1.2).

Besides driving safety, aspects pertaining to user experience such as usability, aesthetics or emotions are critical determinants for the success of human-machine interfaces (HMIs) (Norman 2013, p. 10). Consequently, designers of automation interfaces for on-road vehicles must not neglect hedonic user requirements such as the practicability of NDRTs – a major distinguishing feature compared with aviation automation interfaces that are tailored to professional, well-trained users. Addressing this, uncertainty displays for passenger vehicles should be designed with both driving safety and user experience in mind. Peripheral awareness displays are a promising means for achieving this. In contrast to digital instrument cluster displays, peripheral awareness displays do not require focal attention to perceive information (McCrickard, Chewar et al. 2003; Matthews, Hsieh and Mankoff 2009). Rather, users can remain aware of system uncertainties while attending an NDRT with focused attention. While the display cues are unobtrusive for the majority of time, users are notified of significant changes through perceptible increases in salience. Several publications have highlighted the usefulness of peripheral awareness displays in the context of driving (Loecken, Heuten and Boll 2015; Borojeni, Chuang et al. 2016; Borojeni, Ali et al. 2016), however the existing body of research lacks an implementation regarding uncertainty communication. RQs 2–4 were formulated to address this (see Section 1.2).

Metaphorically speaking, uncertainty communication renders automated systems more transparent, thus allowing users to gauge their reliance on automation and develop an improved system understanding (Balfe, Sharples and Wilson 2018). At one extreme, an interface could convey detailed technical information about the current state of all subsystems. By this means, users would have access to highly detailed, complex information that could guide them to specific automation insufficiencies and allow a prediction of subsystem failures. On the other extreme, the automated system could derive an overall metric from all individual subsystem states and present it to the user as a system-wide indicator for uncertainty. While this simplification may conceal significant information from users, reducing the amount of available data facilitates a fast analysis of current system capabilities. The functional detail of uncertainty information therefore constitutes a trade-off between increased system awareness and operator overload. Current research within the automotive domain lacks an investigation of this trade-off. RQs 5–6 were formulated to address this (see Section 1.2).

To benefit users, the system uncertainties must not be presented like alarms, but rather as fluent transitions (Yang et al. 2017). Consequently, sudden spikes in uncertainty must be avoided. In addition to inherent system uncertainties relating to probability distributions in neural networks or malfunctioning sensors (Gal 2016), upcoming events or obstacles that might affect system uncertainties such as broken down vehicles, construction sites or unmapped areas can be made available via vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication (Eriksson and Stanton 2017; Kim, Park and Huh 2016). Thus, feedback and feedforward information can be combined to achieve an anticipatory uncertainty communication mechanism that allows for more gradual changes.

Irrespective of the positive effects and the technical feasibility of uncertainty communication, its practical implementation depends on the commitment of automotive original equipment manufacturers (OEMs). Why would carmakers deliberately admit – and even actively convey – insufficiencies of their own systems? Recent analyses of vehicles equipped with automated driving systems tested on US roads highlighted that more than half of all automation disengagements are caused by system failures (Favarò, Eurich and Nader 2018). Despite focusing on pre-production vehicles, the analysis emphasised that current systems are still fault-prone. Considering that *safety* and *ethics* are key determinants for the acceptance of automated driving systems (Liljamo, Liimatainen and Pöllänen 2018), OEMs are challenged to a transparent approach that enables users to assess the true capabilities of automated driving systems. Given that the recent emissions cheating scandal caused a major crisis for the whole automotive industry (West 2019) and led to the loss of trust in automotive OEMs (Bratzel 2018), the necessity for a confidence-building approach is further highlighted. Besides ethical and safety-related reasons, OEMs might choose to convey uncertainties to express their *knowledge of the unknown*. While all OEMs are facing similar challenges, those who are able to assess their own capabilities demonstrate a superior system understanding. Alternatively, government agencies could require OEMs to implement uncertainty displays as a safety feature, similar to, for instance, regulations to tackle blind spots in heavy goods vehicles (HGVs) (Summerskill et al. 2019).

It is the role of this research project – and related efforts – to highlight features that affect human-automation interaction and provide guidelines for lawmakers and professionals to improve the safety and experience associated with automated driving systems.

1.2 Objectives and Structure

This thesis aims to enhance user safety and experience through uncertainty communication in the context of automated driving. Specifically, it targets the previously presented shortcomings associated with uncertainty communication and proposes a novel concept for human-automation interaction. The following RQs are addressed in this research project (see Table 1.1 for the derived experimental research questions and analysed dependent variables):

RQ 1: What are the implications of dynamic uncertainty communication on key parameters of human-automation interaction?

RQ 2: How can the systematic evaluation of peripheral awareness displays be ensured?

RQ 3: How can uncertainty information be intuitively conveyed using peripheral awareness displays?

RQ 4: What are the implications of peripheral uncertainty communication on key parameters of human-automation interaction relative to the use of digital displays?

RQ 5: What are the qualitative and quantitative implications of increasing the functional detail of uncertainty information?

RQ 6: What visualisation layouts and variables are best suited for conveying uncertainties with a higher functional detail?

Figure 1.1 illustrates how each chapter of this thesis relates to the formulated research questions. Chapters 2 and 3 provide the theoretical underpinning of this project. Whereas the former focuses on thematic discussions of related work, the latter aims to establish a valid methodology for addressing the research questions.

Specifically, the literature review outlined in Chapter 2 describes relevant engineering psychology concepts such as human information processing and introduces criteria for evaluating system designs with respect to usability and user experience. Starting with a global definition and classification of driving automation, the HF challenges that were briefly outlined in Section 1.1 are increasingly specified. As such, the *automation conundrum* (Endsley 2017) is discussed and contributing factors are identified, whereby a particular emphasis is placed on trust calibration. Increasing transparency is proposed as a means for overcoming the automation conundrum and existing approaches to uncertainty communication are presented. The chapter concludes with the derivation of the previously formulated research questions.

Chapter 3 discusses the research methodology applied in this research project. As such, an overview of the chosen approach is given and key decisions in regard to sampling and the use of a driving simulator are explained. Further, quantitative and qualitative data collection and analysis methods are discussed under consideration of related metrics. Finally, the deliberations regarding experimental design and statistical methods are outlined.

Chapters 4 to 9 each address the research questions and delineate the experimental work conducted as part of this research project.

Chapter 4 addresses RQ 1 and presents the results of a driving simulator study involving 34 participants. Using a between-subjects design, the implications of dynamically conveying system uncertainties on key human-automation interaction aspects such as trust and workload were evaluated. Table 1.1 summarises how each global research question was

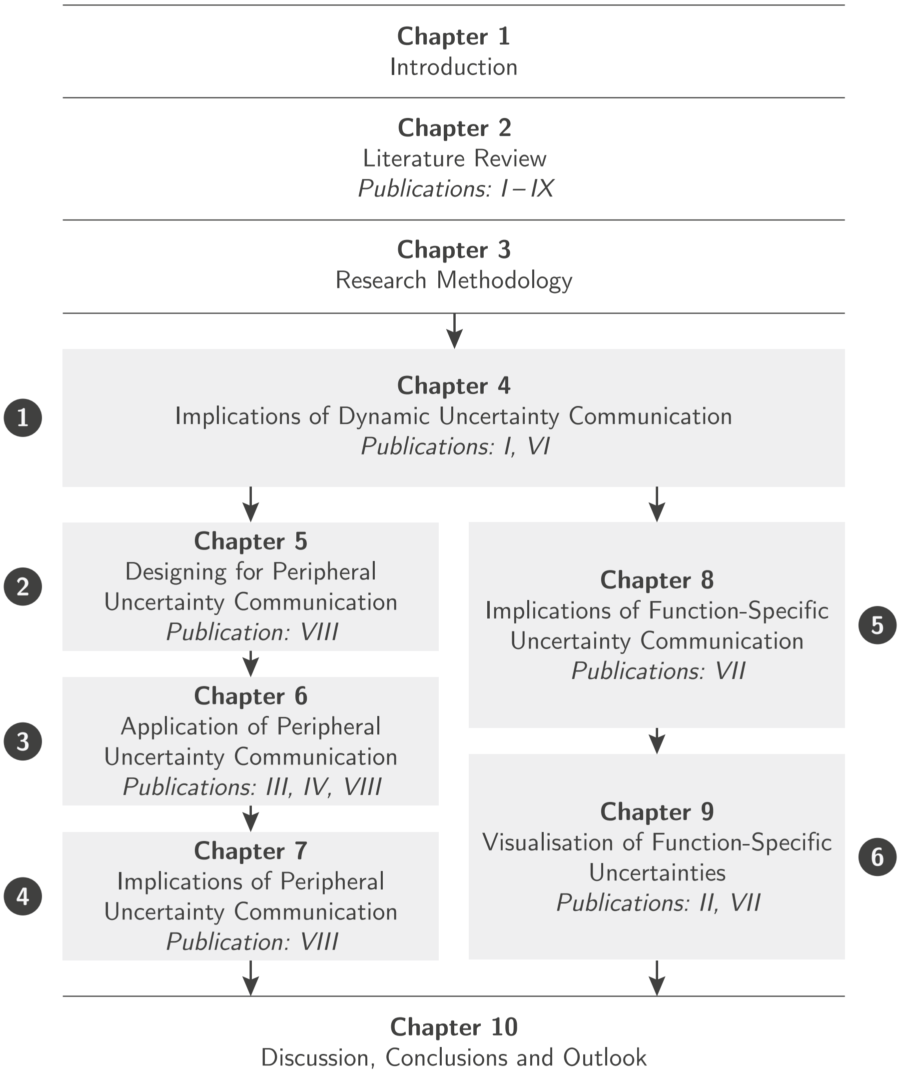


Figure 1.1: Thesis structure, bullets indicating addressed research questions

translated into a series of experimental research questions and presents the dependent variables that were investigated. The experiment outlined in this chapter highlights a series of implications regarding uncertainty communication that build the foundation for subsequent research efforts.

Following Chapter 4, the thesis can be split into two research tracks. Chapters 5 to 7 address the communication of uncertainties using peripheral awareness displays. Chapters 8 and 9 investigate the functional detail of uncertainty information. The thesis concludes with an overall discussion of the findings and summarises the contributions of this research project to the field of HF in human-automation interaction.

Table 1.1: Overview of the overall and experimental research questions and the associated dependent variables

#	Research question	Experimental research question	Dependent variables
1	What are the implications of dynamic uncertainty communication on key parameters of human-automation interaction?	How does dynamically communicating automation uncertainties affect operator trust, SA, and monitoring behaviour?	Single-item subjective trust, responses to SAGAT, monitoring frequency, monitoring ratio, NDRT task performance
		How does operator gaze behaviour relate to trust?	Monitoring frequency, NDRT performance, single-item subjective trust
		Does a higher resolution of uncertainty information affect operator trust and monitoring behaviour?	Monitoring frequency, NDRT performance, single-item subjective trust
		How does dynamically communicating automation uncertainties affect takeover performance and quality?	MTTC, TTT, maximum lateral acceleration
		How does dynamically communicating automation uncertainties affect operator workload?	NASA-TLX responses
		How does dynamically communicating automation uncertainties affect physiological measures?	Heart rate
2	How can the systematic evaluation of peripheral awareness displays be ensured?	What are relevant criteria, dimensions, and methods for the design and evaluation of peripheral awareness displays?	Systematic literature review
		How can these criteria, dimensions and methods be implemented into a comprehensive design process for peripheral awareness displays?	Systematic literature review

Table 1.1: (Continuation) Overview of the overall and experimental research questions and the associated dependent variables

#	Research question	Experimental research question	Dependent variables
3	How can uncertainty information be intuitively conveyed using peripheral awareness displays?	Which light variables afford users to confidently distinguish between several levels?	Fuzzy membership functions
		What light variables are most logical for conveying uncertainties in the context of automated driving?	Questionnaire responses
		What light variables are perceived as most interrupting?	Questionnaire responses
		Which vibro-tactile variable instances clearly convey an increase in uncertainty?	Questionnaire responses
		Which vibro-tactile variable instances clearly convey a decrease in uncertainty?	Questionnaire responses
4	What are the implications of peripheral uncertainty communication on key parameters of human-automation interaction relative to the use of digital displays?	How does the display type affect operator trust and monitoring behaviour?	Single-item subjective trust, monitoring ratio, NDRT task performance
		How does the display type affect takeover performance and quality?	MTTC, TTT, maximum lateral acceleration
		How does the display type affect operator workload?	NASA-TLX responses
		How does the display type affect physiological measures?	Heart rate, HRV, respiratory rate

Table 1.1: (Continuation) Overview of the overall and experimental research questions and the associated dependent variables

#	Research question	Experimental research question	Dependent variables
5	What are the qualitative and quantitative implications of increasing the functional detail of uncertainty information?	How does takeover performance differ when uncertainty is communicated in a function-specific manner compared with an overall system indication?	MTTC, TTT, maximum lateral acceleration
		How do system failures regarding the lateral or longitudinal component of the driving task affect takeover performance under consideration of the functional detail of uncertainty information?	MTTC, TTT, maximum lateral acceleration
		How does the communication of uncertainties with a higher functional detail compare with the presentation of the overall system uncertainties in terms of preference and what are the reasons for preferring one over the other?	Subjective preference, interview responses
6	What visualisation layouts and variables are best suited for conveying uncertainties with a higher functional detail?	Which visual variables are best suited for conveying a guessable order for changes in uncertainty?	Order consistency
		Which visual variables are best received for the purpose of conveying uncertainties?	Subjective preference
		How do preference scores vary with respect to each investigated (driving) function?	Subjective preference

CHAPTER 2

Literature Review

This chapter presents a critical review of related work and provides background information concerning underlying concepts relevant for this research project. The review culminates in the identification and discussion of research opportunities which served as blueprints for the formulation of research questions. Aspects of this chapter, in part verbatim, are published in Kunze, Summerskill et al. (2017, 2018a,b,c, 2019a,b,c).

2.1 Objectives, Scope and Structure

The literature review aims at defining the research problem and providing a comprehensive scientific foundation that allows for the identification of problem areas with further research opportunities. It targets to derive the opportunities immediately from the most significant problem areas and to support their significance with a firm theoretical underpinning.

The review is structured as illustrated in Figure 2.1. Starting with general introductions to key concepts, the analysis becomes increasingly specific and eventually results in pinpointed research opportunities and questions.

At first, the backdrop against which this research project is set will be outlined. This includes a definition of the concepts uncertainty, usability, and user experience (see Section 2.3) as well as an introduction to human information processing, whereby a special focus is placed on the role of attention in perception (see Section 2.4). This lays the groundwork on which subsequent sections are built. Next, the specific application of this engineering psychology project, driving automation, is introduced (see Section 2.5) by defining its key characteristics, describing the tasks of human operators in suchlike vehicles, outlining the technology used to achieve driving automation, and analysing classifications of automated systems. The last-mentioned analysis results in the selection of a taxonomy to be used in this research project. The inherent HF issues of using this taxonomy and the evolutional approach to driving automation are subsequently discussed and further specified for the most significant level of automation (see Section 2.6). The challenges are narrowed down starting with an overarching problem, the *automation conundrum*, and concluding in specific reasons for the HF problems. One of the most significant culprits, inappropriate trust, is further investigated in Section 2.7 with the objective of identifying a potential means for counteracting inappropriate trust and thereby the broader HF challenges. The

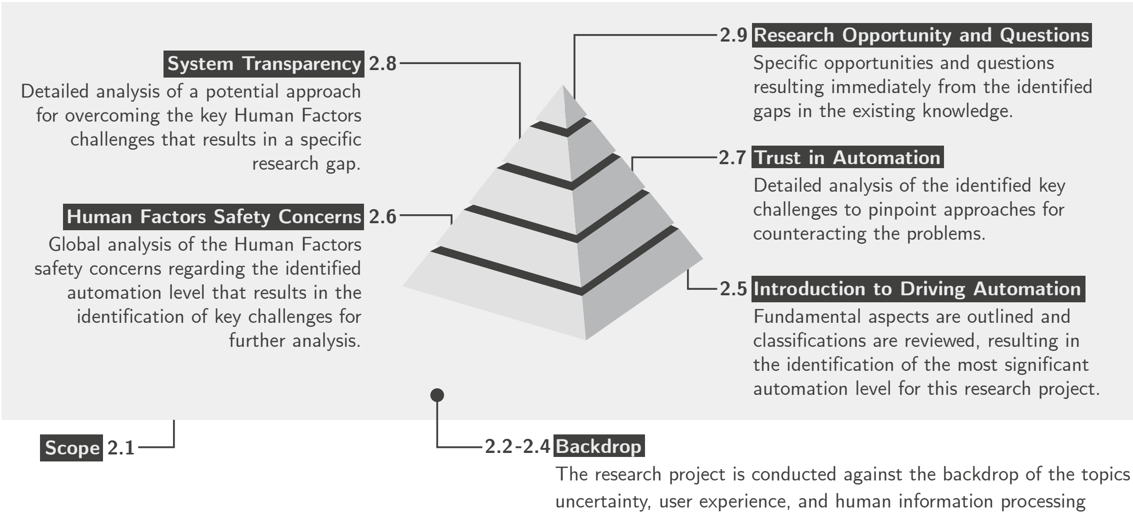


Figure 2.1: Schematic overview of the literature review structure

identified measure – increasing system transparency – is analysed in-depth in Section 2.8. The chapter culminates in specific research opportunities and questions that form the basis for subsequent chapters (see Section 2.9).

This chapter also introduces a colour coding model that will be applied throughout this thesis. Unless explicitly stated otherwise, the colour teal (●) is used to represent system-related content, coral (●) is used to convey information relating to the user, and ochre (●) is applied to attributes relating to the environment and usage context.

2.2 Uncertainty

The measurement of natural phenomena typically results in a distribution of data points around an average value. To account for this spread of values, it is good practice to report a measure of uncertainty alongside the average. For normally distributed phenomena, the mean would be accompanied by the variance or standard deviation of the data set. Increasing dispersion of data points results in decreasing validity and, consequently, uncertainty concerning the average.

The term uncertainty can therefore be understood as a statistical concept that delineates the level of confidence associated with an outcome variable. The impact of uncertainties is thereby not limited to the observation of natural phenomena but extends to any other statistical application such as the approximation of real-world systems with mathematical models (Kennedy and O'Hagan 2001). In this context it is often distinguished between *aleatoric* and *epistemic* uncertainty. Aleatoric uncertainties are induced by random, irreducible influences whereas epistemic uncertainties result from limited knowledge and/or data

(Kiureghian and Ditlevsen 2009; Zaman and Mahadevan 2017). Another differentiation can be made along the three levels of situation awareness (SA) as defined by Endsley (1995b) (see Subsection 2.6.2.2): data uncertainty, comprehension uncertainty, and projection uncertainty (Endsley and Jones 2012). The first level relates to uncertainties in the available data concerning the state of the environment. This may include measurement errors as a consequence of defective sensors. Level 2 refers to uncertainties induced during the interpretation of the data, e. g. a software algorithm that fuses the data of multiple sensors to a combined output. The third uncertainty level concerns uncertainties in projecting future states, for example a model predicting the location of a certain object.

Figure 2.2 depicts the principal sources of uncertainties in complex systems, such as those employed in the context of driving automation. Specifically, uncertainties are induced during data acquisition, data transformation, and output generation. For instance, the occlusion of sensors may impact their reliability and affect acquisition uncertainty. Further errors may be induced by sampling during data transformation. Eventually, mechanical properties such as actuator tolerances may affect the output generation.

This thesis recognises uncertainty as a system parameter that is independent of the user and distinguishes between the uncertainty of the overall system, henceforth *system-wide uncertainty*, and the uncertainty of the system regarding specific functions, hereafter *function-specific uncertainty*. The system-wide uncertainty is defined as a time-specific indicator for the overall reliability and confidence of the automated driving system. In contrast, the function-specific uncertainty delineates the reliability and confidence of individual dynamic driving task (DDT) functions (see Subsection 2.5.1).

In Section 2.8, the benefits of conveying system uncertainties to operators of automated driving systems are outlined based on a review of experimental investigations within the automotive and other domains.

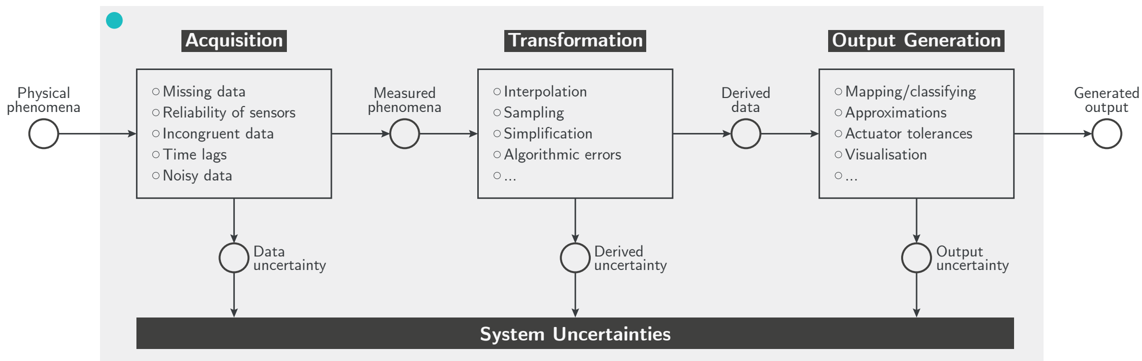


Figure 2.2: Sources of system uncertainties (based on Endsley and Jones 2012; Kiureghian and Ditlevsen 2009; Pang, Wittenbrink and Lodha 1997; Uggirala et al. 2004)

2.3 Designing Enjoyable User Experiences

Traditionally, HF research has focused on factors pertaining to task performance and safety (e. g. Wickens 2002; Parasuraman, Molloy and Singh 1993; Endsley and Kiris 1995). In recent years, however, hedonic and emotional aspects of human-machine interaction gained in relevance (Pettersson et al. 2018).

Whereas the search term `"user experience"` returned approximately 23,800 results for the publication year 2010 on Google Scholar, this number increased to 50,700 in 2018. In contrast, the term `usability` returned 45,400 results for 2018 – 38,300 less compared with 2010 (last update: 28 June 2019). Considering that higher levels of driving automation afford operators more flexibility to engage in NDRTs (see Subsection 2.5.2), the relevance of user experience in the context of driving can be expected to increase further in upcoming years. This section introduces the scope and criteria of user experience compared to those of usability and presents a framework that guides the design of enjoyable experiences.

2.3.1 Definition and Relevance

First, the terms usability and user experience are defined and described in detail. This includes an extraction of criteria pertaining to each term and a delimitation to distinguish between usability and user experience.

2.3.1.1 Usability

ISO 9241-11:2018 (ISO 2018, p. 2) defines usability in the context of human-system interaction as the 'extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use' (see Figure 2.3). Thereby, effectiveness, efficiency, and satisfaction can be

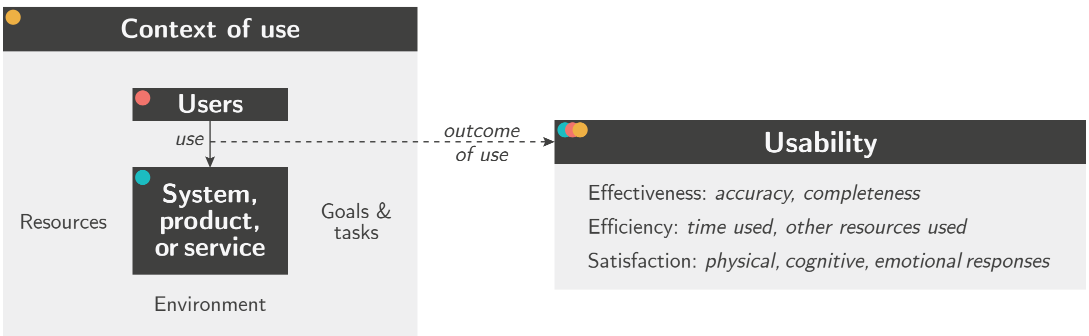


Figure 2.3: Usability as the outcome of usage of a system, product, or service in a specified context of use, adapted from ISO 9241-11:2018 (ISO 2018)

considered as principal usability criteria. Effectiveness describes how accurately and to what extent users achieve the intended outcomes. Efficiency places the used resources such as time or materials in relation to the outcome. Satisfaction describes the extent to which users' expectations are met by the emotional, cognitive, and physical responses to the usage of the product, system, or service.

With a focus on interactive software systems, Seffah et al. (2006) complement the aforementioned criteria and combined multiple usability models into a consolidated Quality in Use Integrated Measurement (QUIM) model that includes the following factors:

- productivity: useful task output relative to resource expenditure
- learnability: the ease with which users can master the features of the system
- safety: the extent to which the system design reduces the risks of harm
- trustfulness: the degree to which a system can be considered trustworthy
- accessibility: the extent to which a system accommodates users with disabilities
- universality: the extent to which a system caters to a wide range of users
- usefulness: the degree of practical utility of a system for the user

Thereby, the factors productivity and efficiency are closely related, as both consider the generated output in relation to the used resources. However, productivity relates to useful output whereas efficiency considers unproductive tasks as well. Thus, users may complete many tasks in a short time (= efficiency), but not necessarily with a higher useful output (= productivity).

While all criteria depend, to some degree, on the abilities of intended users, a lack in accessibility and/or universality would inhibit the proper use of some or all features of a system and would therefore impede the measurement of other criteria such as learnability. Users that cannot perceive any of the display components will not be able to use them. As such, accessibility and universality can be seen as overarching criteria. To evaluate the criteria, the designed system can be assessed against acknowledged standards, for instance ISO 9241-20:2008 (ISO 2006), to ensure that it meets the requirements of a wide range of users under consideration of potential disabilities.

2.3.1.2 User Experience

User experience can be defined as a 'person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service (ISO 2010)'. As such it encompasses both subjective aspects such as emotions and objective facets including, for instance, physiological responses. Further, the definition highlights that changes in user experience are not confined to the use of the product or system. Rather, the time span extends to well before and after usage, being a consequence of not only system performance but also, for instance, brand perception and product presentation.

User experience implies that not only successful, i. e. effective and efficient, task performance but also the pleasant use of a product or system determines the user's judgement. Indeed, usability in its broader sense, as defined in ISO 9241-11:2018 (ISO 2018, p. 2), considers emotional and perceptual aspects that are generally ascribed to the term user experience. Yet, the described usability criteria mainly focus on system-related qualities and only consider emotional reactions as part of the satisfaction criterion.

Therefore, user experience is a broader concept than usability, the latter mainly focusing on the effective and efficient task performance (see Subsection 2.3.1.1).

2.3.2 Components of User Experience

Existing models of user experience subdivide the construct into several core components. This includes the hedonic and pragmatic product attributes as well as users' emotional responses (Thüring and Mahlke 2007; Hassenzahl, Diefenbach and Göritz 2010; Hassenzahl 2005, 2006). Pragmatic product characteristics are related to task performance, similar to the presented usability criteria (see Subsection 2.3.1.1). As such, pragmatic characteristics support users with accomplishing *do-goals*, e. g. writing an email. Contrasting this, hedonic attributes concern the well-being and longing for need satisfaction of users and serve *be-goals*, for instance an identification with brand and product.

In a series of three experiments, Thüring and Mahlke (2007) investigated three principal components of the user experience compound and their interrelations:

1. Perception of instrumental qualities
2. Perception of non-instrumental qualities
3. Emotional reactions to system behaviour

Concluding, Thüring and Mahlke (2007) present the Components of User Experience (CUE) model to summarise the interrelation between the three investigated principal user experience components (see Figure 2.4). As such, user attributes, system features and the usage context form the interaction characteristics that, for their part, affect the perception of instrumental and non-instrumental qualities. Thereby, instrumental qualities delineate characteristic usability aspects such as the ease of use. Non-instrumental qualities include aesthetics and regard the 'look and feel of the system (Thüring and Mahlke 2007, p. 263)'.

Both perceived qualities impact the emotional response of users, such as pleasure and curiosity for well-designed systems and frustration for ill-designed systems. Together, the three components of user experience have an effect on the user's appraisal of the system.

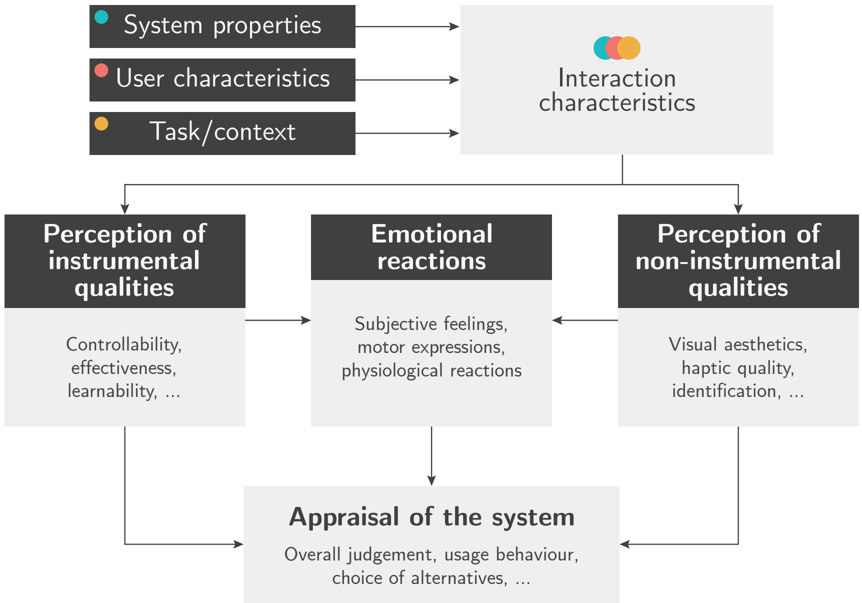


Figure 2.4: The CUE model (adapted from Thüring and Mahlke 2007)

The general framework of the CUE model (see Figure 2.4) and the International Organization for Standardization (ISO) usability model (see Figure 2.3) emphasise that, apart from system properties, the usage context and user characteristics determine what and how users perceive a system. Thus, for any evaluation of either user experience or usability, the usage context and user characteristics must be taken into account.

Chapters 5 and 6 outline how the usability and user experience models were integrated into the development and evaluation of interfaces used within this research project.

2.4 Cognitive Approach to Enjoyable and Safe Designs

Understanding how humans take in, process, and use information is a prerequisite for designing safe and enjoyable experiences. Within the field of cognitive psychology, humans are regarded as information processing systems consisting of several individual, interconnected stages – not unlike those present in contemporary computers (Proctor and Vu 2012).

In this section, the cognitive processes and concepts that condition (multi)task performance in human-machine systems are introduced to identify bottlenecks and human limitations that might impede user experience and safety. The findings inform subsequent design decisions, particularly relating to the selection of suitable modalities (see, e. g., Chapter 6).

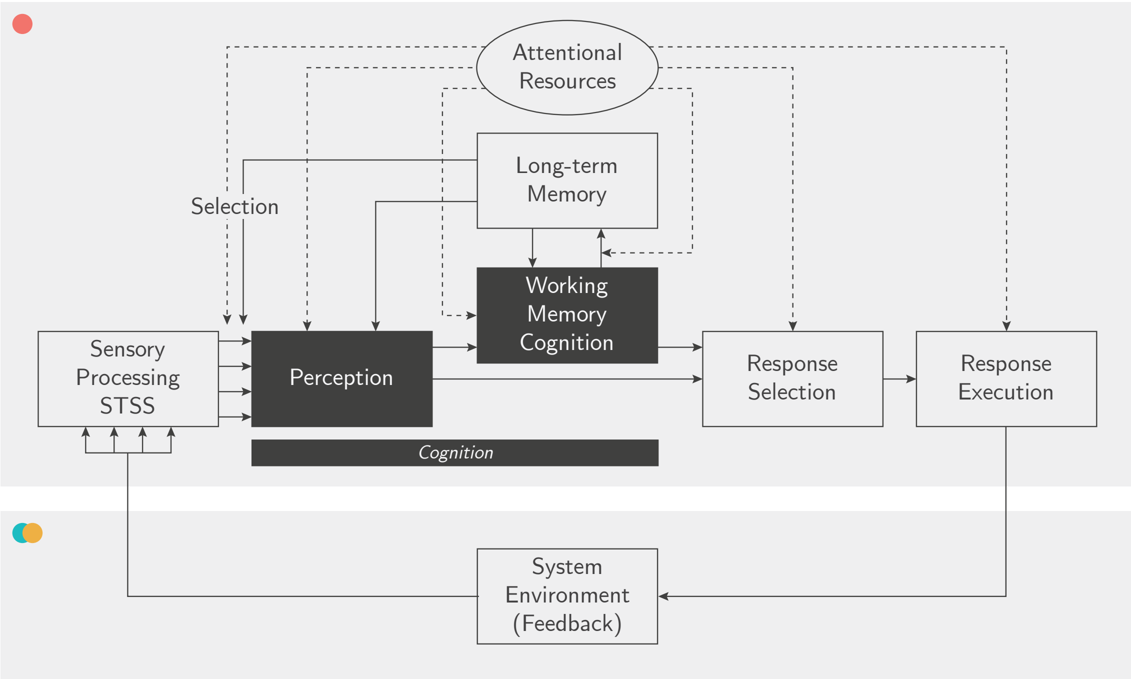


Figure 2.5: Stages of human information processing (adapted from Wickens, Hollands et al. 2013, p. 4)

2.4.1 Stages of Human Information Processing

A model of human information processing stages (see Figure 2.5) provides a useful foundation for the characterisation of mental operations that are typically performed whenever humans complete a task. Environmental stimuli are initially observed by our senses and may then be passed on to the short term sensory store (STSS), where they are typically held for less than a second (Wickens, Hollands et al. 2013). The human senses are commonly structured into five classes of channels (National Research Council (US) Committee on Vision 1985; Winner and Hakuli 2012, p. 344):

1. visual channel
   a) focal vision
   b) peripheral vision
2. auditive channel
3. haptic channel
   a) tactile channel
   b) kinaesthetic-vestibular channel
4. olfactory channel
5. gustatory channel

This *sensation* does, however, not directly translate into *perception*. Rather, only a small ratio of the sensual input will be perceived. Perception comprises the establishment of what the sensual signals mean for an individual. This meaning originates in former experiences, stored in the long-term memory. The perceived signals can subsequently prompt an immediate response, are further processed using working memory, or both. The former follows a direct path from perception via *response selection* to *response execution* (see Figure 2.5). This involves the selection of a response from a multitude of possible reactions and an action upon that selection (Wickens, Hollands et al. 2013). Instead of – or additionally to – the triggering of an immediate response, the perceived signal may be temporarily stored in *working memory* to further think about and learn from it. This process of interpreting the sensed information is referred to as *cognition* and encompasses both perception and working memory (see Figure 2.5). The cognition process can be executed rapidly for simple tasks, such as responding to a perceived red traffic light by braking, or at a slower pace for more demanding activities, for instance the interpretation of an unfamiliar argument presented in a political discussion. In the latter example, this information can be passed on to long-term memory for future debates (Wickens, Hollands et al. 2013). A further distinction can be made between analytic and intuitive cognition, whereby the former entails consciously recollectable, i. e. explicit, knowledge while the latter involves knowledge that is not consciously available, i. e. implicit knowledge, but allows faster responses that are independent from working memory (Patterson and Eggleston 2017; Patterson 2017). System designs can support intuitive cognition by presenting familiar perceptual cues. Consequently, interfaces developed in connection with this research project aimed to employ stimuli that would likely invoke intuitive cognition. To explore this, the intuitiveness of various competing visual and vibro-tactile variables was investigated in the experiments outlined in Chapters 6 and 9.

The pace at which sensory signals are processed does not only change with the task at hand and the intuitiveness of presented cues but is also affected by the sensory channel that

Table 2.1: Typical characteristics of relevant sensory channels (Winner and Hakuli 2012, p. 345)

	Characteristics		
Sensory channel	Alternative naming	Information rate	Perception dwell time
Visual channel	Sight	very high	fast
Auditive channel	Hearing	medium	medium
Tactile channel	Touch	low	very fast
Kinaesthetic-vestibular channel	Kinesthesia	low	very fast

is being addressed (see Table 2.1). The time taken from sensory input until perception, referred to as *dwell time*, varies with the sensory channel. Further, the *information rate*, i. e. quantity of sensed information in a certain time frame (e. g. bits/s), can serve as a measure for the complexity of information that can be conveyed using each channel. As such, signals sensed by the tactile channel are faster processed than those addressing the visual channel. However, the visual channel affords a higher information rate and is therefore more suitable for complex information (Winner and Hakuli 2012). This differentiation is relevant for selecting appropriate modalities within the intended usage context (see, for instance, Subsection 6.1.1).

In most cases, the last stage of human information processing, response execution, leads to a change in the state of the environment and/or system the user is interacting with. This new state creates additional sensory input, thus creating a feedback loop (see Figure 2.5). Importantly, the (limited) availability of *attentional resources* impacts all of the presented stages of human information processing. Attention thereby takes on two separate roles. On one hand, it serves as a filter of information when selecting which sensations are to be attended further. For instance, commuters may completely focus their attention on reading a newspaper and thereby filter out the surrounding noise and activities of the passengers around them. On the other hand, attention serves as a *fuel* and supplies all human information processing stages with mental resources (see Subsection 2.4.3) (Wickens and McCarley 2008; as cited in Wickens, Hollands et al. 2013).

However useful the model presented in Figure 2.5 is in establishing a framework and visualising the human information processing stages, the displayed sequence and clearly differentiated stages are not fully congruent with the associated brain processes (Wickens and Carswell 2012). As such, a task is, for instance, not necessarily triggered by a sensation but can also originate in long-term memory. Nonetheless, the model highlights crucial bottlenecks in information processing and clarifies the significance of attention for perception, an aspect that will be revisited in more detail in a following subsection.

2.4.2 Multitasking

Often people do not perform a single task but rather multiple tasks at once – walking and talking, singing and dancing, or, regrettably, texting and driving. Three mechanisms affect the task performance in a multi-task environment: *resource demand*, *resource multiplicity*, and *resource allocation*. Resource demand refers to the resources that are invested to perform each task, depending on practice and task complexity. More practised and easy concurrent tasks will have less interference than novel, complex tasks. Resource multiplicity highlights the existence of several resource *pools*. Concurrent tasks that address different resource pools

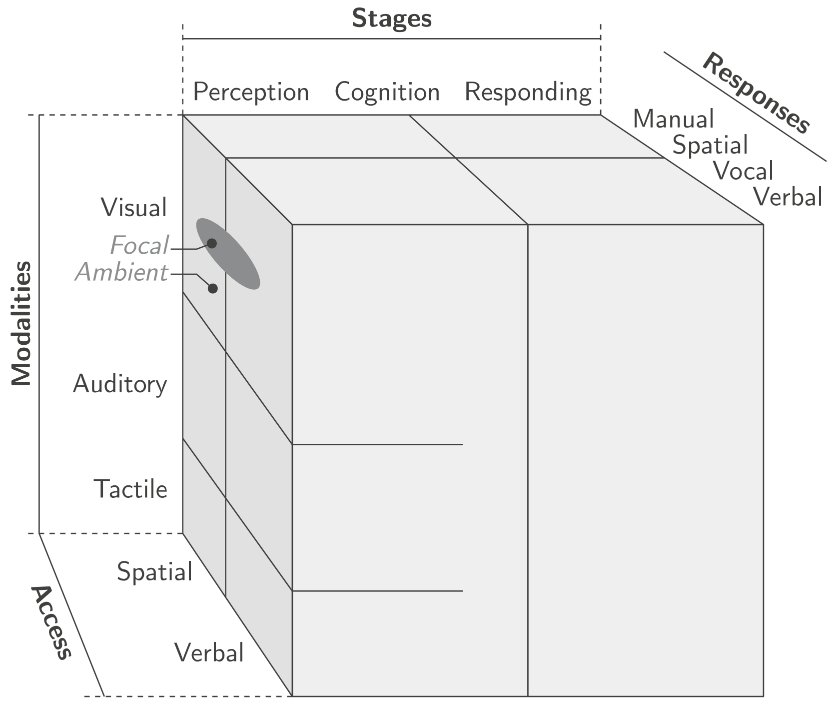


Figure 2.6: Three-dimensional structure of the multiple resource model (adapted from Wickens, Hollands et al. 2013, p. 330)

show less interference than those addressing the same resource pool. Figure 2.6 illustrates the multiplicity of resources, whereby distinct areas on the three-dimensional cubic surface indicate different pools. As conveyed by the cubic structure, three resource dimensions exist. The stages of the human information process (see Subsection 2.4.1) represent one dimension. Cognition and perception address the same pool, whereas response selection and execution resort to a different resource pool. Another dimension is comprised of differences in processing. As such, analogue/spatial processing and categorical/symbolic, e. g. verbal, processing address different resource pools. Thirdly, the perceptual modalities (see Table 2.1) make up the final dimension, whereby a further distinction can be made between focal (mostly foveal) and ambient (mostly peripheral) vision. Tasks that address two different pools lead to a better (faster, more accurate, etc.) performance than those that address the same resource pool (Wickens, Hollands et al. 2013).

The previous two mechanisms, resource demand and resource multiplicity, may lead to a decrement in performance for one or more of the performed tasks due to limited resources. The final mechanism, resource allocation, determines how this decrement is distributed among the tasks, i. e. which tasks have priority. The *multiple resource model* shows several implications for the development of user interfaces. First, the simultaneous execution of two or more novel and complex tasks should be avoided. Second, concurrently executed tasks should access different resource pools. Chapter 6 draws on these aspects and provides an example of how this can be implemented in the context of uncertainty communication.

In addition to performance decrements as a consequence of overloading, the attentional fixation on a particular channel of information or task may lead to breakdowns in task management, referred to as *attentional tunneling* (Wickens 2005). As such, the fixation on one channel or task may lead to the neglect of another. Thus, systems should employ a means for breaking this attentional lock if the situation requires it (Self et al. 2008). The interface designed for the experiment outlined in Chapter 7, for instance, utilises vibro-tactile stimuli to achieve this.

2.4.3 Attention and Perception

Attention is critical for the selection of sensory input that is to be perceived. The concept of attention is commonly explained using the metaphor of an electric torch beam in a dark environment. Similar to how the beam illuminates only a small portion of the environment as it moves along, attention is selectively placed on different areas of interest (AOIs) over time. AOIs are distinct locations in the environment that hold some kind of information relevant for the user. Illuminating only single objects is similar to *focused* attention, whereby the individual is focusing on a single AOI. In contrast, illuminating a larger part of the environment likely renders several objects visible, representing *divided* attention. Illuminating the environment for an extended period of time is similar to *sustained* attention (Wickens, Hollands et al. 2013).

In the context of driving, *visual* attention is particularly significant. Some estimations even place the ratio of visual information input during driving at 90 % (Hills 1980).

2.4.3.1 Visual Attention

The processes delineated by the term visual attention include 'a set of cognitive operations that mediate the selection of relevant [information] and the filtering out of irrelevant information from cluttered visual scenes (McMains and Kastner 2009)'. As a consequence, visual attention determines what we see (Ware 2012). The AOI that is to be attended is determined by the factors *salience*, *effort*, *expectancy*, and *value*, known as the SEEV model. Salience describes to what degree an AOI is distinguishable from its surroundings. The more salient an AOI, the more attention is placed on it. Additionally, the effort it takes to divert attention from one AOI to another influences the attention allocation. Figure 2.7 shows the information access effort (IAE) as a function of the spatial separation between two AOIs. Least effort is required for AOIs that are separated by less than four degrees of visual angle, as attention can be reallocated without eye movement. Up to approximately

20°, eye movements are sufficient to reach neighbouring AOIs, referred to as the *eye field*. Higher visual angles require neck rotations (*head field*), before a full body movement is required from approximately 90° separation. Aside from salience and effort, the degree to which we expect changes in certain AOIs determine which are attended to. Operators of highly reliable automated systems, for instance, do not expect system malfunctions and may therefore fail to monitor suchlike automated systems, an effect referred to as *complacency* (see Subsection 2.6.2.2). Finally, attention allocation is influenced by the *value*, i. e. importance, users place on certain tasks or AOIs (Wickens, Hollands et al. 2013).

The four described factors directly relate to two differing attentional processes. While salience and effort influence the *bottom-up attention allocation*, value and expectancy affect *top-down attention*. Bottom-up attention, also referred to as exogenous attention, is an externally induced process determined by external stimuli (Connor, Egeth and Yantis 2004; Corbetta and Shulman 2002; Itti et al. 2001). Top-down, endogenous attention is based on conscious plans, prior knowledge and present goals (Katsuki and Constantinidis 2014).

The experiment discussed in Chapter 4 highlights the negative effects of employing displays that exhibit high levels of IAE without noticeable changes in salience. Contrasting this, Chapters 6 and 7 present a display that reduces IAE through optimised positioning in the vehicle cockpit and utilises changes in salience to affect bottom-up attention.

The factor expectancy of the SEEV model is of further relevance as humans are particularly poor at noticing unexpected events. Phenomena such as *change blindness* (Healey and Enns 2012) and *inattentional blindness* (Most and Simons 1999) highlight that users may miss even salient changes within their direct field of view. The change blindness phenomenon stresses that users may fail to perceive changes in the environment, particularly if they occur during eye saccades, blinks, or other disruptions to vision that mask natural visual transients. Inattentional blindness, closely related to change blindness, is the failure to notice an object even though you are looking directly at it, i. e. *looking but not seeing*. This effect has been demonstrated in a study that received extensive media coverage: Participants were shown a video sequence picturing a series of actors passing basketballs around and were asked to count the number of passes. At some point during the video, an actor dressed in a gorilla costume walked to the centre of the screen, stopped, beat on his chest, and moved out of the frame again. More than 50 % failed to notice the gorilla (Simons and Chabris 1999). These phenomena emphasise that, without attention, signals are not consciously perceived (Wickens, Hollands et al. 2013). Further, the described phenomena highlight that display designs should not solely rely on the noticeability of changes in the visual field. Chapter 6 presents a multimodal interface to counteract this.

To direct attention to a certain AOI, individuals can either change their gaze, i. e. *overt* attention, or focus attention on a peripheral location without moving one's eyes, i. e. *covert*

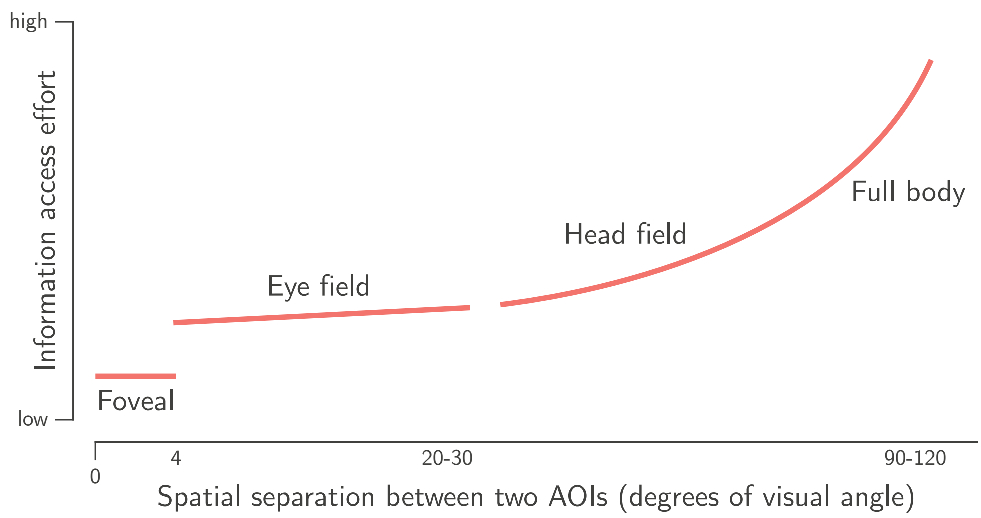


Figure 2.7: IAE as a function of the visual angle separation between two AOIs (adapted from Wickens, Hollands et al. 2013, p. 51)

attention (Posner 1980). Due to the sequential nature of eye movements, overt attention is deployed in a serial manner. In contrast, covert attention can be directed at several locations simultaneously (Bao et al. 2018; Carrasco 2011).

A limited set of visual properties can be detected with a single glance via low-level visual processes, referred to as *preattentive properties*. The term *preattentive* solely refers to the speed with which these properties are perceived (less than 250 ms (Healey and Enns 2012, p. 1171)), and does not indicate that these properties are identified without attention – which would be a wrong conclusion. Among these preattentive visual properties are motion, orientation, density, size, hue, and luminance (Healey and Enns 1999; Huber and Healey 2005; Julesz and Bergen 1983; Treisman and Gelade 1980; Treisman and Gormican 1988). The preattentive processing is supported through the redundant presence of preattentive properties, e. g. the use of hue and size to distinguish features. The parallel, non-redundant presence of properties hinders preattentive processing (Ware 2013). Further, the preattentive properties of visual features are not equally strong but vary in their popout capabilities, with motion, contrast, colour, and size having the strongest effects (Ware 2013). The interfaces employed in Chapters 6 to 9 make use of preattentive properties to reduce cognitive workload and promote faster responses.

2.4.3.2 Human Vision

The previous subsection highlighted that without attention there is no perception. Even more essential, however, is the notion that without sensation, there is no environmental feature the attention can be focused on.

The way we sense visual information with our eyes is not unlike the functionality of a modern digital single lense mirrorless camera. The eye's transparent *cornea* refracts the incoming light much like the optical elements of a fixed focal length camera lens. Similar to the adjustment of the aperture via a diaphragm to regulate the incoming light, the eye's *iris* controls the size of the pupil. The light then passes through a further optical element, the *lens*, which, in contrast to the cornea, is flexible and allows an adjustment of focus – much like the focusing lens (or lens group) of a camera. The adjustment of the eye's focus is called *accommodation*. Eventually, the light reaches the innermost layer of the eye, the *retina*, which converts the light into chemical energy, thus activating nerves that transfer information to the brain. Using the analogy of a camera, the retina would be the digital sensor (Davson and Perkins 2018).

In contrast to a camera sensor, however, the accurate, i. e. sharp, vision is only possible within the *fovea*, a small region of the retina making up only about 2° of the visual field. The fovea is densely packed with *cones*, photoreceptor cells activated at high light levels (photopic vision) that confer colour vision and enable a high visual acuity. The concentration of cones within the retina diminishes with increasing distance from the retina. *Rods*, the highly light sensitive counterpart of cones that enable achromatic vision under low light conditions and are very sensitive to motion, are mainly located in the surrounding region of the fovea, the parafovea. The ratio between rods and cones throughout the retina is approximately five to one (Davson and Perkins 2018).

Due to the limitation of the human eye to only allow accurate vision in the fovea, rapid eye movements, *saccades*, followed by stationary eye periods, *fixations*, are necessary to see detailed information of larger regions (Healey and Enns 2012). Attention enables these rapid eye movements by filtering out visual backgrounds and thus allowing the programming of eye movements based solely on selected signals (Kowler 2009).

The part of vision outside of the fovea is referred to as *peripheral* vision, often further split into near, mid (>30°) and far peripheral vision (>60°) (Sardegna et al. 2002; Strasburger, Rentschler and Jüttner 2011; Simpson 2017). Due to the distribution of rods and cones in the retina, peripheral vision is highly sensitive to motion. Further, peripheral vision is more acute than foveal vision when it comes to the perception of a general scene (Juola 2016), e. g. if you are located in a natural or an artificial environment. Particularly the area within an angle of 30° from the fovea is thereby relevant for driving (Lachenmayr 2006).

Each interface design employed within this research project was informed by the characteristics of human vision. As users were not expected to constantly attend the interface with focal attention and thus sense its information with foveal vision, transitions to higher uncertainty levels were consistently conveyed through animations to allow the perception of changes with peripheral vision.

2.5 Definition and Classification of Driving Automation

The previous sections established the backdrop against which this research project is set. Hereafter, the specific application domain is increasingly specified. As such, this section provides an introduction to the broad topic of driving automation and limits the scope of this thesis within that topic.

First, essential terms such as driving automation are defined to establish a common ground for further analysis (see Subsection 2.5.1). Next, the characteristic tasks of human users in on-road vehicles are discussed in order to emphasise on key changes that the automation of the driving task entails (see Subsection 2.5.2). These changes are further analysed in Subsection 2.5.3, where the interaction as well as interfaces between human operators and (automated) machines are discussed.

In Subsection 2.5.4, the technical aspects of driving automation are outlined to identify the data that is collected by suchlike vehicles and can be made accessible to the user via the HMI. Subsequently, existing classifications of driving automation systems are reviewed (see Subsection 2.5.5) in order to identify gaps in the knowledge which this research will address (see Subsection 2.5.6).

2.5.1 Dynamic Driving Task as Entity of Automation

In general, automation can be defined as 'the use or introduction of automatic equipment in a manufacturing or other process or facility (Oxford Dictionaries 2016a)' to fulfil tasks that would otherwise be performed by humans (Wickens and Hollands 2000). Relating to vehicles, automation can encompass functions such as driving, lighting, or air conditioning. This research project focuses on driving automation which can be defined as the system-controlled performance of at least parts of the DDT on a sustained basis, i. e. irrespective of changes in the driving environment that require a response (SAE International 2018). This definition excludes temporary interventions such as the intermittent brake impulses controlled via an anti-lock braking system (ABS).

The DDT can broadly be considered as 'the sum of processes, which are required for operating a vehicle in on-road traffic (Altendorf et al. 2016, p. 266)' and is a proper subset of the complete driving task.

Donges (1982) characterises the driving task using a three-layered model consisting of navigation, guidance, and control. An alternative model proposed by Michon (1985) adopts the three levels but labels them strategical, tactical, and operational driver efforts. Löper et al. (2006) further specify the driving task by splitting up the guidance level into manoeuvre and trajectory planning. Irrespective of the chosen layer model, the DDT excludes the uppermost level of the driving task. As such, the DDT does not include strategic (Michon

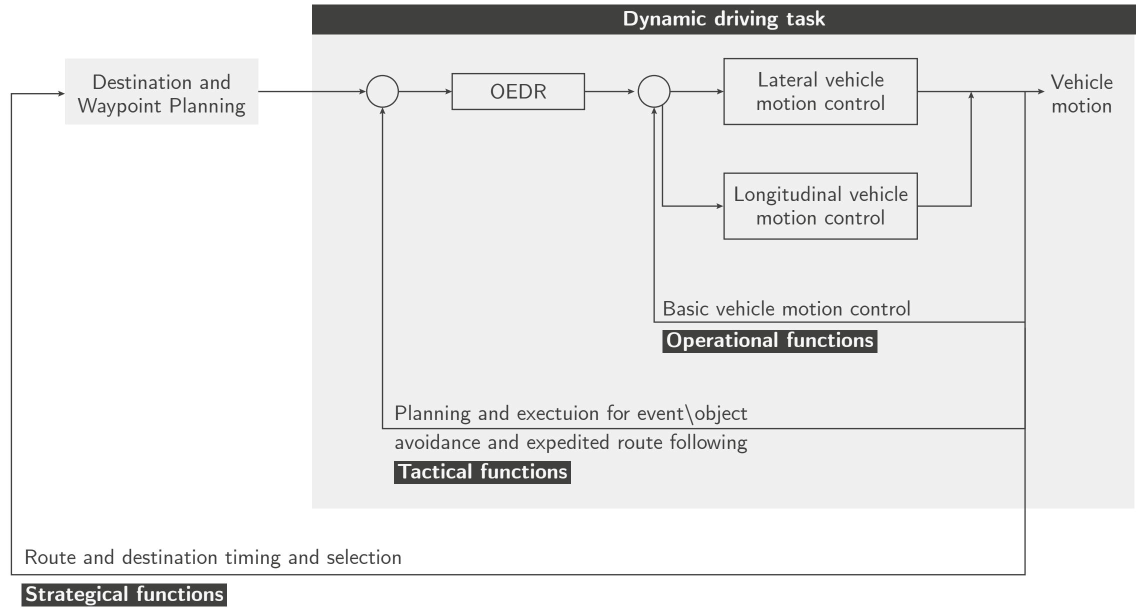


Figure 2.8: Schematic view of the driving task (adapted from SAE International 2018)

1985) or navigational (Donges 1982) efforts, such as trip planning. Rather, it involves the planning of manoeuvres as well as trajectories (e. g. overtaking) and the intuitive, natural reactions, e. g. steering to keep in the lane. Furthermore, the monitoring of the environment and responses to changes, summarised as object and event detection and response (OEDR), are also part of the DDT (SAE International 2018). Figure 2.8 shows a schematic overview that defines the scope of the DDT (tactical and operational functions) as a proper subset of the complete driving task (strategical, tactical, and operational functions).

The popularity of driving automation systems with the public has introduced several terms intended to imply an identical meaning to the one that has just been presented, rendering a disambiguation necessary. The terms *autonomous* and *self-driving* are often used as alternatives to driving automation. However, both are criticised by the Society of Automotive Engineers (SAE) as potentially confusing and misleading (SAE International 2018). Autonomous infers the general independence of outside entities, which is not necessarily the case in all scenarios. Self-driving, on the other hand, is ambiguous regarding the existence of a human operator in the vehicle (SAE International 2018).

Besides, SAE International (2018) recommends against using the term *automated vehicle*. As with self-driving, it fails to distinguish between vehicles that are designed to have a human operator and those that are not. Additionally, the term does not specifically relate to driving. This thesis follows the demonstrated recommendations and considers driving – and not the vehicle – the entity of automation.

2.5.2 User Tasks in Automated Driving

Traditionally, the different activities of human operators in manually controlled vehicles have been divided into two – primary and secondary – or three – primary, secondary, and tertiary – different classes.

The primary task comprises the whole DDT as described in Subsection 2.5.1. When using the two-class model, the secondary task includes all activities that differ from the DDT (Wierwille 1993). In case of the trisected model, secondary tasks refer to activities that are directly related to the primary task and aim at either informing other road users or at responding to environmental changes. Examples are the use of indicators when changing lanes or the engagement of the fog tail light on nebulous roads. Tertiary tasks encompass all other activities, for example the manipulation of the infotainment system controls (Bubb et al. 2015). With increasingly automated driving systems and consequently less engagement of the operator in the driving task, the formerly tertiary tasks may become the most significant activities, rendering them the *primary* tasks in automated driving. Thus, the previously employed nomenclature promotes confusion and needs to be adapted.

In HF research regarding automation, the following terms are frequently used and will be adopted within this research project (Pfleging and Schmidt 2015):

- Driving-related tasks (DRTs): activities related to the DDT, i. e. a combination of the primary task and the secondary task of the trisected model
- Non-driving-related tasks (NDRTs): activities not related to the DDT, i. e. formerly the tertiary tasks in the three-class model

The consideration of NDRTs is significant for the design of automated driving systems, with researchers even arguing that the acceptance of vehicles equipped with automated driving systems depends on the degree to which the comfortable engagement in NDRTs is possible (Miglani, Diels and Terken 2016).

Several studies have investigated the relation between increasing automation and the exertion of NDRTs as well as the impact of NDRTs on safe driving performance. Results indicate that higher automation levels lead to an increased engagement in NDRTs. In a simulator study that examined the relation between automation and task-unrelated thoughts of pilots in aircraft cockpits, Casner and Schooler (2014, p. 433) found that higher levels of automation afford pilots time to think ahead but lead to more task-unrelated thoughts if there is no interaction with the system.

Even without vehicles with higher automation levels available, drivers frequently perform NDRTs such as texting while driving (Carsten et al. 2012; Fitch et al. 2013; Klauer, Dingus et al. 2006; Norman 2015, p. 84; Klauer, Guo et al. 2014). Carsten et al. (2012, p. 747) investigated this in a study that tested the involvement of drivers in NDRTs. The results show that the engagement in distracting tasks is likely to increase with the degree of automation.

Jamson et al. (2013, p. 116) came to a similar conclusion, but documented an adaptivity of the behaviour depending upon the traffic situation. Additionally, Naujoks, Purucker and Neukum (2016) observed an increase in NDRT engagement for partial and highly automated driving (SAE levels 2 and 3, see Subsection 2.5.5.2). Supplementing this, Vollrath, Schleicher and Gelau (2011, p. 1134) did not observe a higher engagement in NDRTs for lower levels of automation, i. e. cruise control (CC) or adaptive cruise control (ACC) (SAE level 1).

Lin, Ma and Zhang (2018) interviewed 20 drivers of Tesla cars equipped with an *Autopilot* feature, corresponding to SAE Level 2 of driving automation (see Subsection 2.5.5.2). Universal among all participants was the increased engagement in NDRTs while the driving automation system was activated. In a longitudinal study, Large, Burnett, Salanitri et al. (2019) observed that participants interacting with an SAE level 3 automated driving system readily engaged in visually demanding NDRTs when given the opportunity. Further, engagement in NDRTs increased from already high levels on day one to the fifth and last day of the experiment.

In a driving simulator study involving 30 participants, Mok et al. (2017) investigated how much time drivers engaged in NDRTs need to take over the driving task following a TOR. The results indicate that a transitional period of 5 s to 8 s is required for safe takeovers.

Interestingly, the engagement in NDRTs – and thereby distraction from the DDT – is not necessarily impeding the safe driving performance in vehicles equipped with automated driving systems. In fact, Miller, Sun et al. (2015) showed that participants engaged in NDRTs are less likely to display signs of drowsiness than those monitoring the performance of the automated system. Further, the results indicate that driving performance following a takeover is not impaired by the previous engagement in NDRTs. In reference to Yerkes-Dodson law (Yerkes and Dodson 1908), the higher levels of arousal relative to the passive monitoring of the system have likely contributed to the improved takeover performance.

Moreover, the type of NDRT affects takeover performance. As such, visual-manual NDRTs lead to larger performance decrements than auditory-vocal NDRTs (Wandtner, Schömig and Schmidt 2018).

The analysed research strongly indicates that NDRTs are a key aspect to be considered in the design of automated driving systems. It can be assumed that users will increasingly engage in NDRTs with higher levels of driving automation (SAE level 2+) and are less likely to actively monitor the performance of the system if not supported. Further, NDRTs offer a chance in that they prevent the user from becoming drowsy and may consequentially even enhance the manual driving performance following takeovers.

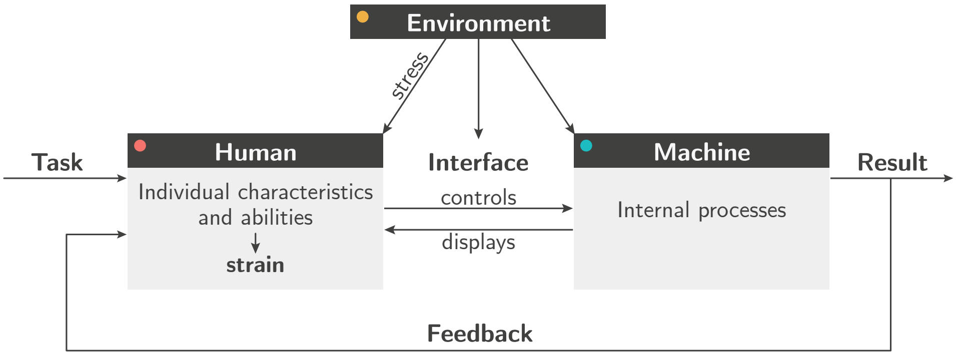


Figure 2.9: Schematic feedback loop of human-machine systems (adapted from Bubb et al. 2015, p. 18)

2.5.3 Human-Machine Interfaces for Driving Automation

Now that both the entity of automation and user tasks in the context of automation have been established, it remains to be addressed how humans interact with automated driving systems. Human-machine interfaces (HMIs) are integral elements of human-machine systems that determine this interaction. As illustrated in Figure 2.9, the structure of human-machine systems resembles a feedback loop: Human and machine, which are integrated in a system, perform a task with a certain result. Both task and result, the feedback, are known and can be compared with each other, thus closing the loop. Regarding on-road vehicles, the superordinate task is the safe performance of the driving task that results in specific driving parameters, e. g. lane deviation, which simultaneously serve as feedback for the human-machine system.

Within human-machine systems, HMIs are any points of interaction between human and machine where information is transferred from the human operator to the machine or vice versa (Bubb et al. 2015). Human operators accomplish this information transfer via *controls* that generate an input to the machine. Machines convey information by producing an output which is perceivable by humans (see Subsection 2.5.3.3). This thesis uses the general term *displays* as the superordinate designation for technical elements that convey information from the machine to the user, i. e. the counterparts of controls (see Subsection 2.5.3.3).

Human operators, machines, and the interaction between them are affected by environmental factors, such as lighting. These factors may represent a *stress* on humans, who – depending on their individual characteristics and abilities – react to them in a particular way, known as *strain*. During investigations of HMIs, the stress on participants must be controlled to minimise unwanted effects on the interaction.

With increasing automation, the relationship between human and machine changes. Where humans used to manually perform every aspect of the driving task, automated systems can aid human activities or completely replace them. This poses novel questions regarding the control, responsibility and authority in human-machine systems (Flemisch, Heesen et al. 2012). In the following, two concepts for the interaction between human operators and (partially) automated machines are presented, namely *cooperative* and *adaptive* automation. Further, displays for HMIs of automated driving systems are reviewed.

2.5.3.1 Cooperative Guidance and Control as Interaction Concept

The concept of cooperative guidance and control is based around the thesis that human operators and the automation system should share the task of guiding and controlling the vehicle in order to make use of their different abilities (Altendorf et al. 2016, p. 265). Flemisch, Bengler et al. (2014, p. 345) argue that human operators are still superior in understanding contextual aspects while automation systems can contribute faster reaction times. They differentiate the cooperative activities via the levels of the driving task (see Subsection 2.5.1 and Figure 2.8): navigation, manoeuvre guidance, trajectory guidance, and control. Two practical implementations of the concept are presented in the following.

Conduct-by-Wire

The human-machine interaction concept Conduct-by-Wire (CbW) distributes the levels of the driving task in a static, hierarchical manner, with the driver being responsible for manoeuvre commands and the automation system for trajectory guidance and control. In practice, the human driver selects from a range of executable manoeuvres that are presented in an HMI, for example a touch interface. The technical feasibility of the concept has been successfully proven but the driver acceptance is yet unknown (Flemisch, Bengler et al. 2014, pp. 347–351; Kauer, Schreiber and Bruder 2010, p. 1220).

H-Mode

An alternative concept to CbW is Horse-mode (H-mode) which transfers the *horse metaphor* to automated driving systems and is more dynamic than CbW. As with riding a horse, the driver can allocate more influence to the automation system, i. e. *loose rein*, or can decrease its influence, i. e. *tight rein*. Automation system and driver are coupled via haptically active interfaces that provide feedback, e. g. an active side stick. The faster the grip of the human operator, the tighter the rein, i. e. the more control is transferred to the driver (Flemisch, Bengler et al. 2014, pp. 351–356). Similar to CbW, the system suggests executable manoeuvres via a visual HMI. Additionally, the suggestions are also communicated via a haptic interface, e. g. by exerting a force on an active stick. The manoeuvre is selected through the simplified movement of the inceptor, e. g. moving an active stick to the right/left for a lane change. Pilot studies have revealed that the concept is capable of lowering the workload compared to manual driving while also allowing the driver to perform NDRTs (Flemisch, Kelsch et al. 2008).

Other Implementations

Walch et al. (2016) proposed an alternative approach to cooperative guidance and control. Instead of aiming for cooperation in all situations, they analysed the use of cooperative interfaces in situations that require the driver, i. e. when the automation reaches its operating limits. To investigate this, a simulator study was conducted in which the interaction types (touch and speech input) and performance were assessed. The results show that the great majority of participants preferred the selection of manoeuvres to the takeover of fully manual control. However, the study was based on the assumption that the system is able to warn human operators 200 m before an obstacle, allowing 18 s for the decision-making process. Studies that investigated the takeover time, e. g. Gold, Damböck et al. (2013), used intervals of 7 s. It is therefore questionable if the results also remain valid in situations that require faster decisions.

Critique of Cooperative Guidance and Control

Including drivers in high-level decision-making tasks requires them to have a basic understanding of the surrounding scene. The low-level guidance and control tasks can be left to the system in order to reduce the workload of the driver (Kauer, Schreiber and Bruder 2010, p. 1221). Depending on the frequency of decisions, this may lead to long periods of passive monitoring – which has been found to be a high workload activity (Hancock 2013; Grubb et al. 1994). On the other hand, drivers might experience additional workload in situations that require very frequent input, such as traffic in city centres, as users have to perceive and understand the possible manoeuvres and select the most appropriate one very quickly. Besides, it is questionable if users prefer to be involved in high-level guidance and control when other OEMs offer solutions that afford increased flexibility to engage in NDRTs. Concepts that allow operators to disengage completely from the DDT and only require their input in certain situations might be preferred. Within this research project, the concept is considered as one of various potential ways of interacting in future vehicles to ensure that the proposed solution can be implemented in practice irrespective of the overarching interaction concept.

2.5.3.2 Adaptive Interfaces as Interaction Concept

The variation in the abilities of human operators is a major factor that distinguishes vehicle and aircraft automation out of an HF perspective. Due to differences in, for instance, age and experience, the specific demands for interfaces vary. Besides, different driving conditions may require higher or lower degrees of transparency (Saffarian, Winter and Happee 2012). Adaptive interfaces can account for these variations and adjust the presented information, e. g. in regard to its salience or detail.

Piechulla et al. (2003) investigated an interface that blocks incoming phone calls in high workload situations in order to prevent overloading. The results of a field experiment suggest that the proposed adaptivity can reduce operator workload. Research from Körber, Gold et al. (2016) indicates that the age of drivers does not have a significant influence on the success of takeovers, even though their way of handling situations varies. Interfaces for takeovers that are adaptive to the age of the operator might therefore not be relevant.

When designing adaptive interfaces it is essential to have sufficient knowledge of the users' current needs. Hiding features to reduce workload in highly-demanding situations may have positive implications on task performance, but may negatively affect user experience (Jung, Kaß et al. 2017). Thus, without accurate, time-specific knowledge of the driver's needs, adaptive interfaces should be avoided.

In Chapter 10, the implications of uncertainty communication in regard to both cooperative guidance and control (see Subsection 2.5.3.1) and adaptive interfaces are highlighted.

2.5.3.3 Displays for Human-Machine Interfaces

The paramount purpose of displays is to make relevant data about the environment as well as system functions available to the human operator (Bubb et al. 2015, p. 275). This includes warnings, e. g. a pressure loss in the brake system, or indications, e. g. a notification that the full light beam is engaged. Signifiers that provide information about the current system state are major sources of feedback in human-machine systems and can increase the transparency of the machine's internal processes (see Figure 2.9).

In driving automation systems, displays can also provide information about the intentions of the automation, i. e. feedforward information. Further, rationales for a certain action can be conveyed. Koo et al. (2015) investigated these types of information in a driving simulator study. They examined the effects of presenting information about *how* the vehicle is going to act compared to communicating *why* the system performs a certain action. The results show that information about the reasons for certain actions, e. g. obstacle ahead, is preferred by users and leads to a better driving performance than the presentation of simple feedforward announcements on its own, e. g. the car is breaking. Although a combination of both led to the best driving performance, participants reported it as unpleasant, potentially resulting from information overload.

Also addressing feedforward information, Naujoks, Forster et al. (2017) proposed a visual interface that incorporates the potential benefits of vehicle-to-vehicle communication, e. g. communicating upcoming system limits in advance. Similar to Koo et al. (2015), the designed visual interface conveyed *how*- and *why*-information to the driver. The proposed design received high usability and comprehensibility ratings in a small-scale expert evaluation, affirming the results of Koo et al. (2015).

The findings highlight that displays of automated driving systems should not only convey feedback (see Figure 2.9) but also feedforward information to support human-automation interaction. The communication of changes in the reliability or uncertainty of an automated driving system due to upcoming obstacles or events is an example of feedforward information relevant for this research project.

Irrespective of the type of information conveyed, displays can potentially address all human sensory channels (see Section 2.4). Both the olfactory and the gustatory channel are rarely used in automotive HMIs (Wintersberger, Dmitrenko et al. 2019). Primarily, the visual, auditive, and tactile channel are actively addressed by existing HMIs and are therefore likely familiar to users.

Besides, multiple channels can simultaneously be addressed to make use of their distinct advantages. Interfaces that simultaneously address multiple sensory channels are referred to as multimodal interfaces (Lu et al. 2016). To analyse the state of the art regarding automotive displays, recent publications, particularly conference proceedings, were reviewed and are subsequently discussed, categorised according to the sensory channel they address.

Visual Displays

Visual displays can either address the focal or peripheral vision of the user (see Figure 2.6). Both types are subsequently analysed and recent research is reviewed.

Focal Vision In contrast to auditive and tactile interfaces, focal visual displays afford users to acquire information at their own time while also allowing for the communication of complex elements via graphical representations. Depending upon the degree of abstraction, visual displays can be categorised as digital, analogue, representational, or contact analogue (see Figure 2.10) (Bubb et al. 2015, p. 276). Higher levels of abstraction are thereby thought to increase the cognitive effort required for linking the presented information to the real world. Traditionally, the instrument cluster was the preferred location for visual displays. More recently, head-up displays (HUDs) have become increasingly popular since they do not require users to glance away from the field relevant for driving (FRD) (see Figure A.2).

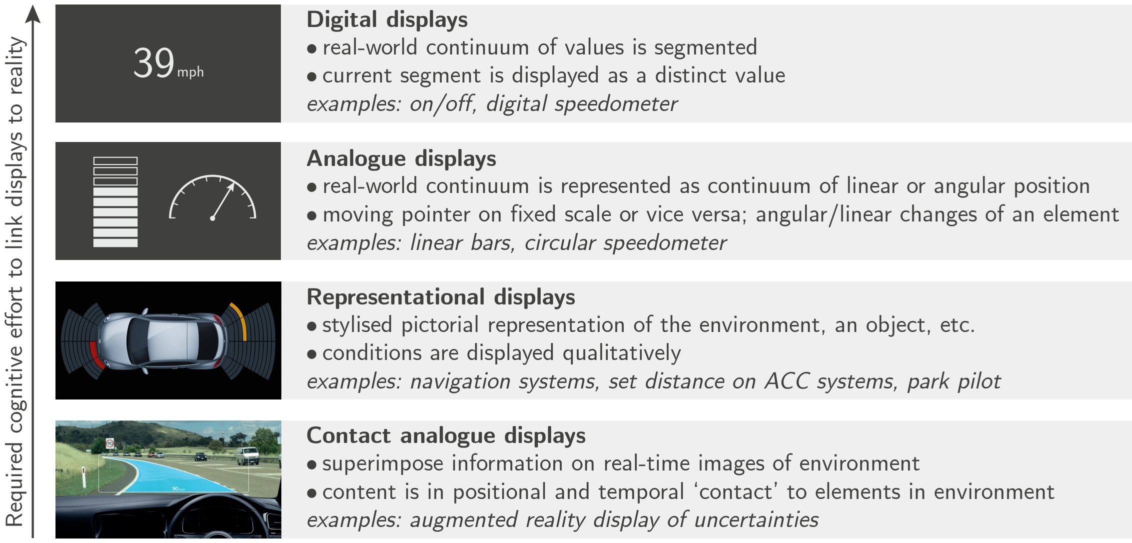


Figure 2.10: Display formats of focal visual displays and the cognitive effort needed to link content to reality (adapted from Israel 2012, p. 18; Bubb et al. 2015)

HUDs project content into the field of view (FOV) of the driver and thereby guide the attention of the user towards the front and thus the FRD (Winkler, Kazazi and Vollrath 2018; NHTSA 1995). Despite this advantage, the use of HUDs could lead to an occlusion of the FRD and promote attentional tunneling – two major challenges of using HUDs (Walch et al. 2016).

The most recent research efforts regarding the use of visual displays in automated driving systems have focused on contact analogue head-up displays (caHUDs) that convey information in positional and temporal *contact* to the real-world environment (see Figure 2.10), also referred to as augmented reality (AR) displays. Regarding manually driven vehicles, research efforts mainly investigated the use of caHUDs for navigational purposes (Pfannmüller et al. 2015; Topliss et al. 2018). In the context of driving automation, Damböck et al. (2012) used a caHUD to communicate information about the vehicle's trajectory, detected objects, and identified road signs with the objective to support the driver's understanding of the scene. Results of a driving simulator study with 24 participants show significantly lower reaction times after automation failures compared to a system without the visual aid. Supplementing this research, Haeuslschmid et al. (2016) also investigated the use of caHUDs to keep human operators aware of their surroundings and presented a framework for information presentation. Their concept separates the display area into several segments, whereby driving-related information was preferred in close proximity to the driving scene.

While the previously presented approaches using HUDs project visual information onto the windshield or transparent plates, Lauber, Böttcher and Butz (2014) investigated the use of head-mounted displays (HMDs) that allow a more flexible FOV. However, HMDs require dedicated, likely user-specific glasses to present the information and are therefore more difficult to integrate in vehicle systems, particularly when content needs to be positioned relative to the vehicle cockpit and driving scene.

Peripheral Vision By definition, the interfaces described above require focal vision. As several other tasks in the context of driving address the same channel (Borojeni, Chuang et al. 2016), the interface must compete for attentional resources. Considering the multiple resource theory proposed by Wickens (2002) (see Section 2.4), this can lead to bottlenecks in critical situations such as takeovers. Displays that address peripheral instead of focal vision access different resources and can consequently reduce workload (Leibowitz et al. 1982).

In an attempt to explore these interfaces further, Loecken, Heuten and Boll (2015) investigated the benefits of ambient lights for lane change manoeuvres in manual driving. Within a driving simulator study, the intensity of an ambient light increased with the suspected uncertainty of the driver and moved from the B-pillar towards the A-pillar as a vehicle approached from the left rear. Their results show that the ambient lighting improved both speed and quality of the human decision-making process.

Also regarding peripheral vision, Borojeni, Chuang et al. (2016) assessed the suitability of ambient lights to announce imminent takeovers and convey contextual information such as the location of an obstacle. The results indicate that ambient lights can decrease reaction times without negatively affecting operator workload. Further, Borojeni, Ali et al. (2016) investigated the use of peripheral cues in the context of automated driving to aid users with the resumption of DRTs after being interrupted by an NDRT, e. g. receiving a phone call while entering an address into the navigation system. Participants that were provided with peripheral cues showed shorter resumption times and less errors. Subsection 6.1.3.1 presents a more extensive literature review that expands upon these findings.

Vocal and Acoustic Displays

In contrast to visual interfaces, acoustic cues completely free the FOV and can be used to guide visual attention. However, the time available for comprehending its meaning can usually not be controlled by the user and auditory displays likely interfere with NDRTs (Ho and Spence 2005).

In a driving simulator study, Naujoks, Forster et al. (2016) compared the suitability of generic auditory cues with that of more natural speech as a means for communicating the intentions of a driving automation system. The results show that participants favour speech as it lowers the visual workload and interferes less with NDRTs, in this case reading. While speech output allows drivers to keep visually engaged in the NDRT, auditory icons require human operators to look up in order to identify the reason for warnings. Their findings, however, only consider visually demanding tasks, i. e. reading. The results are likely different for NDRTs that address the auditory channel, such as conversations with passengers or music. Further research highlights that the selection between abstract and language-based warnings depends on criticality (Politis, Brewster and Pollick 2015b). This suggests that the best-suited type of auditory communication, i. e. language-based or abstract, depends on the usage context, e. g. on the criticality of the conveyed information or presence and type of NDRTs. Overall, auditory cues are primarily beneficial for warnings and notifications.

Vibro-Tactile Displays

Vibro-tactile displays 'consist of arrays of vibrating elements coupled to the skin (Van Erp and Van Veen 2001, p. 99)' and were shown to reduce mental effort relative to visual displays in an automotive context (Van Erp and Van Veen 2001). Tactile cues have similar benefits as auditory cues in that they free the FOV. Additionally, tactile cues are less likely to interfere with NDRTs. On the contrary, tactile stimuli are less suitable for communicating complex information (Ho, Tan and Spence 2005).

As the driver's seat is constantly in contact with the operator, much research has focused on integrating tactile interfaces in vehicle seats. Chang, Hwang and Ji (2011), for instance, evaluated the suitability of vibro-tactile seat feedback for several use cases, including navigation, receiving telephone calls, and a driver drowsiness system. The results indicate that vibro-tactile stimuli reduce response time relative to visual or auditory interfaces.

Schwalk, Kalogerakis and Maier (2015) investigated different forms of encoding information regarding takeovers using vibro-tactile cues in the context automated driving. For this, a tactile seat matrix consisting of 47 vibration motors was used in a driving simulator setting. The results indicate that the transfer of control from driver to automated system is best conveyed by cumulatively activating the vibration motors, starting from the top of the back rest and ending at the front of the seating surface. Conveying a TOR, i. e. change of control from automated system to driver, was rated best for the opposite direction, i. e. cumulative activation beginning from the front of the seating surface and ending at the top of the back rest.

With the objective of maintaining an awareness of the surroundings in the context of automated driving, Telpaz et al. (2015) communicated spatial information of nearby vehicles using a vibro-tactile seat matrix. In line with Chang, Hwang and Ji (2011), results of a driving simulator study suggest that vibro-tactile feedback can reduce reaction times.

Combined, the findings suggest that vibro-tactile cues are particularly useful in situations that require immediate driver reaction.

Multimodal Displays

Multimodal displays were shown to increase the perceived urgency but also annoyance of drivers (Politis, Brewster and Pollick 2013, 2015a) and lead to faster reaction times (Politis, Brewster and Pollick 2014). These findings are supported by Ho, Reed and Spence (2006) who found improved reaction times following audio-tactile warnings compared to unimodal auditory or tactile cues. Further supporting the benefits of multimodal cues, Naujoks, Mai and Neukum (2014) investigated the effects of audio-visual signals compared to unimodal signals and concluded that the former result in faster reaction times. ISO 9241-920:2009 (ISO 2009) also recommends the combination of modalities, particularly to compensate for sensory channels that are overloaded.

Conclusion

The presented recent conference publications highlighted distinct applications of each display type. Visual interfaces using HUDs or HMDs were found to be useful for the communication of complex information that is related to the surroundings, such as planned vehicle manoeuvres. Displays addressing peripheral vision, on the other hand, were more suited to prompt attentional shifts and for conveying warnings, such as vehicles approaching from the rear or imminent takeovers. Auditory displays were applied for notifications and warnings, but are more likely to interrupt NDRTs (see Subsection 6.1.1). Vibro-tactile stimuli and multimodal interfaces were associated with faster reaction times and should consequently be reserved to convey information of high urgency.

2.5.4 Automation Technology

An understanding of the technology that is applied to achieve driving automation is required to identify the input that automated systems can provide for the HMI displays in order to improve driving safety and user experience. In order to automate the DDT, three main tasks must be fulfilled by a system (Fridman 2017):

1. Localisation and mapping: determining the exact position of the vehicle within the environment
2. Scene understanding: identifying the position and movement path of surrounding objects
3. Movement planning: calculating the trajectory to get to the target destination

Table 2.2: Typical characteristics and potentials of sensor technologies (Winner and Hakuli 2012, p. 220)

	Vision	**Lidar**	**Radar**	**Range Imager**
Wave length [m]	$10^{-7} - 10^{-6}$	10^{-6}	$10^{-3} - 10^{-2}$	$10^{-7} - 10^{-6}$
Dependence on weather	yes	yes	low	high
Resolution [# readings]:				
horizontal	$10^2 - 10^3$	$10^2 - 10^3$	$10^1 - 10^2$	$10^1 - 10^2$
vertical	$10^2 - 10^3$	$10^1 - 10^2$	10^1	$10^1 - 10^2$
temporal	$10^1 - 10^5$	10^1	10^1	10^1
Primary measurements:				
position	-	+	+	+
velocity	-	-	+	-
brightness patterns	+	+	-	+
Applications:				
detection of objects	+	+	+	+
identification of objects	+	+	o	o
detection of lanes	+	o	-	-
detection of traffic signs	+	-	-	-

The technology that enables the fulfilment of these functions consists of a combination of sensors, actuators, algorithms, and processors (Schweber 2016). For identifying the information that is available to the system, sensors are of particular importance.

2.5.4.1 Sensors for Driving Automation

Sensors provide the input information that is needed to achieve automated driving. In the following, sensors that are generally employed in vehicles equipped with automated driving systems are presented.

GPS and IMU

The Global Positioning System (GPS) is the primary subsystem that is used to determine the current position of the vehicle and is essential for navigation. By analysing signals from a minimum of four low-orbit satellites, GPS can achieve an accuracy of approximately 1 m (Schweber 2016).

However, GPS is susceptible to signal blockings, e. g. in tunnels, and interference, e. g. by radio waves. As a consequence, the signal can be lost for several minutes, making an inertial guidance system necessary that does not depend on external signals. For this, inertial measurement units (IMUs) are used that can calculate the motion of the vehicle

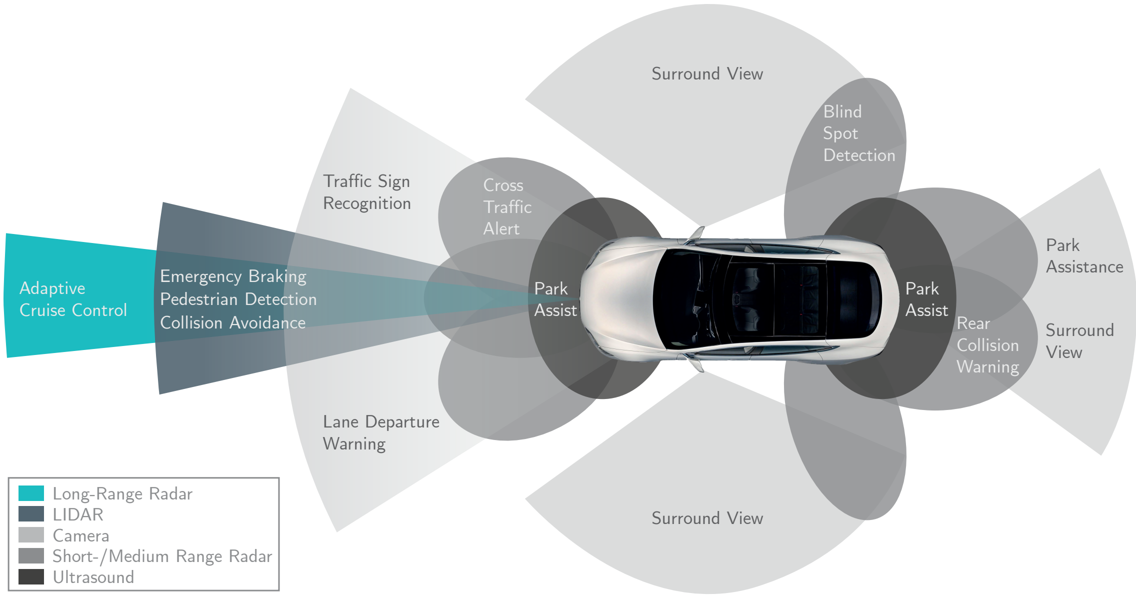


Figure 2.11: Sensors generally employed in current automated driving systems (adapted from Parrish 2015)

and thereby its current location given a starting point provided by GPS. This is achieved through the combination of accelerometers and gyroscopes that provide information about the rotational and linear movement of the vehicle (Schweber 2016).

Ultrasound, Radar, Lidar, Range Imagers and Machine Vision

Driving automation systems require a multitude of sensors to gain an understanding of the surroundings, e. g. detecting lanes or nearby objects. For this purpose, *ultrasound*, *radar*, *lidar*, *range imagers* and *machine vision* are employed (Schweber 2016).

With its first applications in the automotive domain dating back to the 1990s, ultrasonic sensors have mainly been used to support manual parking. The distance to objects is calculated based on the time it takes for a sound wave to travel between object and sensor. Ultrasonic sensors are characterised by inexpensive manufacture, high accuracy as well as relative independence of both weather conditions and materials, but also by a limited range (typically about 150 cm) and relatively long measuring times (limited by velocity of sound). Thus, they are mainly used for the detection of nearby objects (Winner and Hakuli 2012, p. 110).

Radar is an acronym for *radio detection and ranging* and uses radio waves for measuring distances to obstacles. The technology has its origins in the military and was first used as a sensor for ACC in commercial vehicles in 1998. Its main advantages are the robustness against weather conditions, the long range (more than 250 m possible), and the possibility of directly measuring the relative velocity of moving objects by making use of the Doppler effect. The spatial and temporal resolution, however, are limited (see Table 2.2) (Winner and Hakuli 2012, p. 172).

Light detection and ranging, better known in its abbreviated form lidar, is an optical measurement system. The general principle is similar to radar, but instead of radio waves the sensor uses UV, IR, or visible light – resulting in significantly shorter times between the emission and detection of pulses (Winner and Hakuli 2012, p. 172). Lidar compares favourably with radar regarding both resolution and area coverage. It is, however, significantly more susceptible to weather conditions.

Range imagers can capture visual and spacial information with one shot and can assign distance measurements to each pixel. This significantly reduces both required time and processing power for post-editing (Winner and Hakuli 2012, p. 187).

For machine vision, an image sensor maps brightness patterns to multidimensional image signals that are evaluated to gain secondary information, such as the position, movement, and type of objects (Winner and Hakuli 2012, p. 198). Machine vision is considered to be the sensor technology that provides the most comprehensive information and is characterised by the highest resolution (see Table 2.2). A drawback is the necessity of complex post-processing because the primary measurements are limited to brightness patterns. A computational model that is increasingly used for signal processing is deep learning (see Subsection 2.5.4.2).

With infrastructure and traffic being optimised for (human) visual perception, machine vision can, in theory, perceive all information that is required for a sufficient scene understanding (see Table 2.2).

Sensor Fusion

As previously described, each sensor is associated with distinct advantages and disadvantages. While radar, for example, can be used for measuring the longitudinal distance to an object, machine vision can complement this by classifying the detected object. Sensor fusion aims at combining the data of different sensors to reduce the shortcomings of individual sensors. Figure 2.11 shows the sensors and their applications within automated driving systems.

2.5.4.2 Artificial Intelligence

With more readily available graphics processing units (GPUs) and storage space, improved algorithms, specialised software, and more accessible infrastructure, data as well as funding, artificial intelligence (AI) is becoming increasingly capable of signal processing and eliminating the major drawback of machine vision, i. e. complex post-processing (Fridman 2017).

Machine learning (ML) is a type of AI that uses algorithms to analyse and learn from data in order to perform certain tasks instead of hand-coding rigid instructions (Barker 2017). Traditional approaches to ML, however, still required manually entered functions in order to identify specific features in an image that could be an indication for a particular object. These features and their interactions were very difficult to identify and minor changes in the image, such as rotations, could lead to errors. As a consequence, computer vision was highly brittle (see Subsection 2.6.2.1) and could not be used for the identification of a broad range of objects (Barker 2017).

Inspired by the current understanding of neural networks within the human brain, deep learning (DL), a class of ML, does not require hand-crafted features for identifying objects but uses a general-purpose learning procedure instead (LeCun, Bengio and Hinton 2015, p. 436). DL utilises large amounts of unstructured data, e. g. images, to learn hierarchical models consisting of several processing layers that achieve an increasingly abstract representation of the initially raw data. In each layer, the raw input is transformed into a slightly more abstract representation. The layers consist of artificial neurons that each assign weightings to the input and generate a certain output that functions as the input of the next layer. To determine appropriate weightings that lead to correct outputs, the *backpropagation* algorithm is used to calculate gradients for the weightings that indicate a necessary change of the selected parameter (LeCun, Bengio and Hinton 2015; Barker 2017; Copeland 2016).

The output of the final layer consists of probabilities for the detection of a certain object or for the performance of a particular manoeuvre (Copeland 2016). Even though the artificial neural networks can be fooled to falsely classify objects, machine vision using DL has recently achieved 'near-human-level performance (Nguyen, Yosinski and Clune 2015, p. 427)' in object classification. Within this research project, the inherent uncertainty of the neural network, i. e. the resulting probability distributions, can be used as the basis for dynamically communicating the capabilities of the system to driver (Gal 2016).

2.5.5 Current Classifications and Examples

As outlined in the previous section, recent technological advances have rendered full driving automation possible. However, current implementations vary in their capabilities, prompting the definition of different levels of automation. Predefined levels ranging from low (i. e. fully manual) to high (i. e. fully automated) are often used as a framework to describe the automation degree of a system or its parts in regard to specific tasks (Miller 2005;

Parasuraman, Sheridan and Wickens 2000). This is useful for characterising the capabilities and limitations of a system and determining the role of the human user.

Concerning on-road vehicles, the automation degree is often described as a function of the lateral as well as longitudinal support provided by a technical system (see Figure A.1). More comprehensively, the automation degree – or automation levels – are characterised by the degree to which the DDT is fulfilled by a supporting system.

2.5.5.1 Comparison of Existing Classifications

Several government agencies as well as independent organisations have worked on general taxonomies to create a common understanding of automation levels. In 2013, the US National Highway Traffic Safety Administration (NHTSA) and the German Bundesanstalt für Straßenwesen (BASt) were first to publish classifications. Both taxonomies consist of five mutually exclusive levels that range from *no automation* to *full automation*. They differ, however, in terms of language, consistency, multidisciplinarity, and clarity. Being aimed at United States (US) governments, the language of the NHTSA classification is more normative than descriptive and provides a lower degree of both detail and clarity than the BASt classification. Besides, the NHTSA levels apply to vehicles instead of driving automation, which could lead to ambiguities (see Subsection 2.5.1). Furthermore, features and functions such as ABS or electronic stability control (ESC) that do not automate the driving task on a sustained basis but are rather active safety systems are included in the NHTSA taxonomy (SAE International 2018).

In 2014, the SAE published an additional classification that was largely based on the BASt levels. The SAE altered the BASt taxonomy through the addition of another level, the provision of definitions and examples, and the description of a systematic distinction between levels.

Following the publication of the SAE taxonomy, the International Organisation of Motor Vehicle Manufacturers (OICA) also adopted the general BASt classification and included the sixth level proposed by SAE.

2.5.5.2 SAE Taxonomy for Levels of Driving Automation

This thesis uses the classification proposed by SAE as it is currently the most detailed and comprehensive taxonomy, specifically regarding higher automation degrees. Table 2.3 shows the most recent publication of the SAE (J3016_201806) relative to the levels proposed by BASt and NHTSA. The SAE taxonomy distinguishes between six different levels using the following aspects:

- sustained lateral and longitudinal vehicle motion control

- OEDR
- DDT fallback
- operational design domain (ODD)

The first two aspects can be summarised as the DDT (see Subsection 2.5.1 and Figure 2.8). The term DDT fallback is defined as the performance of the DDT or the achievement of a minimal risk condition by either the automated driving system or the user following a takeover. If a driving automation system performs neither the lateral nor the longitudinal vehicle control on a sustained basis, it is to be allocated to level 0. A driving automation system that takes over lateral or longitudinal control is characterised by level 1. ACC, which adjusts the velocity to maintain a set distance from the vehicle ahead, is an example for level 1 driving automation In case both directions are controlled by the system, level 2 is chosen. For levels 0-2, however, the driver still has to monitor the system and is responsible for OEDR and DDT fallback.

With level 3, the driver does not have to monitor the vehicle at all times but must be ready to perform the DDT fallback. For level 4, this fallback is performed by the system itself, e. g. to achieve a minimal risk situation.

A level 5 driving automation system is fully capable of performing all aspects of the DDT as well as the DDT fallback in all situations, i. e. all ODDs. A system at this level is able to operate the vehicle in every situation as human drivers could. The ODD of the systems of the other levels can be characterised by, for example, 'geographic, roadway, environmental, traffic, speed, and/or temporal limitations (SAE International 2018)'.

Limitations of the outlined SAE levels for driving automation must be noted. While helpful for providing a common understanding concerning automation capabilities, the SAE levels do not reflect the allocation of legal responsibilities. In fact, current legislation demands that liability, by default, rests with the driver – independent of the automation level (Mirnig et al. 2019). The SAE classification lacks specifics in order to be used as an insurance blueprint as it leaves room for ambiguities. For instance, the taxonomy is missing a detailed definition of minimal risk conditions. As such, it is currently unclear what manoeuvres qualify for SAE level 3 and level 4 systems. Moreover, it can be considered unlikely that a level 3 system is capable of informing a user of an imminent takeover in a timely manner while not being able to achieve a minimal risk condition itself. Further, the hierarchical level structure may misleadingly suggest an increase in sophistication for higher levels. However, this is not necessarily the case as the higher automation level might only apply to a highly specific ODD, for instance airport shuttles (Law Commission of England and Wales and Scottish Law Commission 2018).

Table 2.3: Automation levels as defined by the SAE in relation to the automation levels of BASt and NHTSA (European Road Transport Research Advisory Council 2015; Gasser and Westhoff 2012; National Highway Traffic Safety Administration 2013; SAE International 2018)

Automation Level		Narrative definition	Sustained lateral and longitudinal vehicle motion control	OEDR	DDT fallback	ODD	BASt level	NHTSA level
Driver peforms 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	Driver only	0
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	Assisted	1
2	**Partial Driving Automation**	The sustained and ODD-specific execution by a driving automation sstem of both the lateral and longitudinal vehicle notion 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	Partially automated	2
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	Limited	Highly automated	3
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	Fully automated	3/4
5	**Full Driving Automation**	The sustained and unconditional (i. e., not ODD-specific) performance by an ADS of the entire DDT and DDT fallback withoutany expectation that a user will respond to a request to intervene.	System	System	System	Unlimited	-	

2.5.6 Conditional Driving Automation as Safety-Critical Level

The classification using levels is well-suited as a common ground for communication across several disciplines that are involved in the realisation of driving automation systems and can prevent confusion (Creaser and Fitch 2015; Lee 2018).

Nevertheless, the presented concept using levels is criticised as being not beneficial for the design process of vehicles that still require the driver to fulfil tasks. On the one hand, the levels could misleadingly suggest that the system can operate at a specific automation degree throughout the whole trip and may lead to false user expectations (Department of Defense 2012). For a similar reason, the German transport minister asked Tesla Motors to stop advertising their level 2 driving automation system as an *Autopilot*, as this may lead to false expectations regarding the capabilities of the system (Hern 2016). Similarly, the Association of British Insurers presses carmakers to avoid using the term *autonomous* as part of their marketing (Cellan-Jones 2018). In contrast to the NHTSA classification, the SAE taxonomy considers limitations in the ODD and can therefore communicate the specific capabilities of the system relative to the driving situations (e. g. highway or city roads).

On the other hand, the generalised levels focus on the capabilities of the technical system and suggest to leave the remaining tasks to the driver (National Research Council 2014; Norman 2015, p. 76; Department of Defense 2012; Lee 2018). This is particularly problematic as recent research findings suggest that the driving performance following takeovers deteriorates as the level of vehicle automation increases (Strand et al. 2014, p. 218).

Especially in regard to partial and conditional automation this results in major problems as users are assigned to tasks which humans are particularly inept for – e. g. long-term monitoring (Norman 2015, pp. 81–84; Sheridan 2002). In both levels, humans are not immediately involved in the manual control of the car, i. e. they are *out of the loop* (Norman 1989, p. 4). While level 2 automation requires human operators to perform the OEDR at all times and thus keeps them involved in the DDT to some degree, vehicles with conditional automation (level 3) completely relieve users from the DDT. When the automated system is not capable of handling a situation, a *conditional driver takeover*, also referred to as *rapid onboarding*, is required and communicated by the system as a TOR (Casner, Hutchins and Norman 2016, p. 75). The fact that the human operator is required to be ready to perform the DDT at all times without previously monitoring the surroundings and performance of the vehicle poses significant HF challenges, which render SAE level 3 the most challenging level of automation out of an HF perspective (Kyriakidis et al. 2017; Casner, Hutchins and Norman 2016; Norman 2015; Trimble et al. 2014).

2.6 Safety Concerns Pertaining to Conditional Driving Automation

Relieving human operators from the DDT and expecting them to resume control in case of situations outside the ODD or system failures induces several safety concerns. Within this section, appropriate sources are selected and analysed to identify challenges that must be considered when designing HMIs for vehicles equipped with automated driving systems. While SAE level 3 was identified as the most safety-critical degree of automation (see Subsection 2.5.6), the safety concerns also more generally apply to all automated systems that (temporarily) remove human operators from the control loop of a task.

2.6.1 Identifying Sources of Information

As of today, no vehicle equipped with conditional driving automation is available to purchase. While Tesla's *Autopilot* and Daimler's *Driver Pilot* can perform the lateral and longitudinal vehicle motion control within a limited ODD, they require the driver to carry out OEDR and can therefore be regarded as level 2 automated driving systems.

The majority of research within the automotive domain relates to lower levels of driving automation that do not allow the driver to completely disengage from the DDT. However, problems experienced with lower levels of automation can likely be transferred to conditional driving automation, such as longer reaction times when using ACC (Vollrath, Schleicher and Gelau 2011).

Aviation

One of the main fields in which theoretical and applied research concerning automation has been conducted is aviation (Meyer 2015, p. 82). Starting with a patent of a gyroscopic stability augmentation device in 1891, the automation of the key aircraft functions 'aviate, navigate, communicate (Billings 1997, p. 66)' was advanced throughout the past century to increase safety, reliability, economy, and comfort (Wiener and Curry 1980).

Up until the 1980s, '[...] the concept of automating as much as possible was considered appropriate'. However, '[. . .] serious questions have arisen and incidents/accidents have occurred which question the underlying assumption that the maximum available automation is always appropriate [. . .] (Air Transport Association of America 1989, pp. 4–5)' – a potential omen for driving automation.

The results of studies in aviation and the automotive domain (Wickens and Kessel 1979; Vollrath, Schleicher and Gelau 2011, p. 1134) suggest that the interaction between a human operator and an automated system entails similar difficulties in aircraft and on-road vehicles (Cummings and Ryan 2013, p. 3). For instance, an activated flight automation system allows pilots to engage in other tasks much like an automated driving system affords users the engagement in NDRTs. Further, takeovers from automated to manual control are a persisting – albeit rare – challenge for pilots and aircraft designers (Trösterer, Meschtscherjakov et al. 2017).

However, there are key differences between the two domains that affect the transferability of identified issues from aviation to automotive applications. Firstly, pilots are selected as a result of an extensive application process, are highly-trained and their aptitude is regularly evaluated. In contrast, drivers have very limited training, usually only at the very beginning, and virtually every adult member of the public can obtain a driver's licence. This leads to a higher variation in both cognitive and physical capabilities of the operators (Creaser and Fitch 2015, p. 83). Moreover, the variability of *sky users* compared to *road users* is limited. Whereas air traffic is confined to strictly-controlled aircraft, road users may include cyclists and pedestrians who might exhibit unpredictable behaviour.

Besides, the required time to respond to emergencies is much shorter in automobiles than it is in aircraft – seconds compared to usually several minutes (Casner, Hutchins and Norman 2016, p. 73; Hoeger et al. 2011, p. 129). As studies have shown that shorter response times negatively affect the effectiveness of feedback, e. g. alarms (Lee, McGehee et al. 2002; Young and Stanton 2007), the efficient communication of information is even more crucial.

Additionally, occurring errors can be handled better in aviation due to the availability of co-pilots, air traffic control, and other supporting aspects (Creaser and Fitch 2015, p. 83). Furthermore, the operators of aircraft have a significantly higher obligation to consistently monitor the system as they are responsible for the lives of potentially hundreds of passengers.

Overall, this leads to higher requirements regarding the simplicity, accessibility, and universality of automated system interfaces in the automotive domain.

2.6.2 The Automation Conundrum

As described in Chapter 1, the challenges regarding the automation of vehicles go beyond technical feasibility. In fact, Neale and Dingus (1998) argue that the most difficult problems regarding driving automation systems are *soft*, i. e. relating to HF issues such as usability, rather than *hard*, i. e. concerning technological issues.

The term *automation conundrum* refers to the combined HF challenges that the automation of systems entails. With automated systems becoming more sophisticated, robust, and reliable, it will be increasingly unlikely that human operators can effectively monitor system performance, be aware of critical information, and, as a consequence, re-engage in the DDT when needed (Endsley 2017, p. 8; Saffarian, Winter and Happee 2012, pp. 2296–2297). In the following, the specific HF challenges are analysed in more detail.

2.6.2.1 Increasing Complexity and System Limitations

In 1977, General Motors (GM) introduced the first passenger vehicle with a programmed microcontroller (Free 1977). Since then, the amount of software in cars has increased exponentially and resulted in approximately 100 millions lines of code per vehicle in 2016 – more than five times as many as in the flight control system of a Boeing 787 (McCandless, Doughty-White and Quick 2015). This is one indicator for the increasing complexity in cars – even without driving automation.

Particularly in aviation, increasing complexity and system limitations due to brittleness and literalism were identified as major challenges (Billings 1997, p. 186). In the following, the specific challenges will be discussed in more detail.

Increasing Complexity

According to Billings (1997, p. 186), high complexity has contributed to several aviation incidents and accidents in which the automation functioned as it was designed to, but the pilots did not understand the intended function and either used the system 'beyond its capabilities or without regard to its constraints'.

The functions an automated aircraft control system can fulfil are accessible via several *modes of operation* that include specific sets of executable instructions suited for different circumstances, e. g. in regard to the vertical control of an aircraft. As more sophisticated technology became available, increasingly complex, mode-rich systems have been designed to improve the flexibility and capability of the automation system (Sarter and Woods 1995, p. 7). This has led to a large variety of available automation modes and a longer delay between user input and system feedback, resulting in novel attentional and cognitive demands for pilots (Woods 1993). In addition to manual flying skills, the characteristics of each mode in different contexts have to be understood and – as a prerequisite – the pilots must be aware of the currently selected mode. Given the multitude of different modes and the possible interactions among them, the chances that the human operators have a sufficient mental model of the system's operation and are aware of all nuances are not high (Davenport and Harris 1992; Wiener 1989). Through eye tracking studies it has been revealed that the automation mode is regularly misremembered and system status displays are ignored (Sarter, Arbor et al. 2007, pp. 355–356).

Particularly when the automation fails the complexity of systems can contribute to serious accidents. In 2009, for example, Air France flight AF 447 ended with the loss of life of all aboard when the plane's pitot pipes malfunctioned and the pilots were not able to correctly understand the problem in time and keep the plane from stalling (BEA 2012). In another aviation accident, unawareness of an activated autopilot led to the death of all passengers and crew (Aircraft Accident Investigation Commission (Ministry of Transport Japan) 1996).

Similar problems can be observed in the automotive domain. In 2016, a Tesla Model S equipped with a level 2 driving automation system collided with a nearby trailer out of a parking position after the remote autopark feature *Summon* was engaged. The feature is designed to park the car at a pre-configured distance to nearby objects after the driver has exited the vehicle. In this case, however, the driver was not aware of the activated system and failed to perform the required monitoring task – which subsequently led to a damaged windshield because the sensors could not detect the obstacle (Clamann and Cummings 2016).

Casner, Hutchins and Norman (2016, p. 75) argue that the complexity is further increasing due to V2V communication and the possibility of joint manoeuvres that happen too fast to be monitored by a human operator.

Overall, the described problems lead to an increased need for mode and situation awareness (see Subsection 2.6.2.2) that results in cognitive demands regarding the following aspects (Sarter and Woods 1995, pp. 17–18):

- the establishment of an accurate mental model concerning system operation
- the tracking of system surroundings and input
- the perception and comprehension of system feedback

Sarter and Woods (1995, p. 6) argue that these novel demands tend to congregate in the most critical situations which would require the support of the automation system the most, thus increasing the workload of the operator (Billings 1997). This has been referred to as *clumsy automation* (Wiener 1989).

As a consequence, the system must provide feedback that supports the human operator with the fulfilment of these demands (Sarter and Woods 1995, p. 18). The appropriate amount and type of feedback will be discussed in Section 2.8.

Brittleness and Literalism

Automated systems are only capable of handling situations they have been programmed for. The aspects *brittleness* and *literalism* are based around the problem that the increasingly complex systems are difficult to design for and test against all possible conditions, particularly with machine learning (see Subsection 2.5.4) (Yoshida 2016). Emphasising this, recent crashes involving two of Boeing's newest 737 Max aircraft can possibly be attributed to a malfunctioning anti-stall system (Young 2019). Brittleness refers to undesired behaviour of the automation system in unanticipated conditions and can lead to unexpected transitions of control from the automation system to the human operator (Endsley 2017, p. 6; Billings 1997, p. 187; Woods and Cook 2006).

Literalism, on the other hand, refers to the constraints and insensitivities of the automated systems when the operator goals are falling outside of conditions that were expected by the designers (Billings 1997, p. 190). Both aspects were identified as major reasons for aviation incidents in the past (Billings 1997, pp. 187–190).

Cummings and Ryan (2013, p. 1) argue that these problems can be observed in automated driving as well, e. g. when Google's automated driving system caused an accident while travelling on a road that was not previously mapped (DeBolt 2011). Additionally, Casner, Hutchins and Norman (2016, p. 72) point out that similar issues have been experienced with GPS navigation in vehicles. The assisting systems function very well in the majority of cases but can lead to accidents when unusual situations are encountered (BBC News 2009). Especially with DL, the decision-making process of the driving automation system is non-transparent, which could potentially lead to unexpected TORs.

Conclusion

Particularly in combination, the increasing complexity and the limitations of automated systems can lead to dangerous situations that could potentially diminish the anticipated safety benefits of driving automation systems. While brittleness and literalism increase the likelihood of unexpected takeover situations, the high system complexity makes it difficult for the user to comprehend the situation in time to prevent dangerous situations or accidents.

2.6.2.2 The Out-of-the-Loop Performance Problem

The out-of-the-loop (OOTL) performance problem refers to the inability of human operators to re-engage in the DDT after a failure of the automation system or the exceedance of its operating limits. As a consequence, users are slow to detect and diagnose problems for effective takeovers (Wickens and Kessel 1979; Young 1969). Endsley and Kiris (1995, p. 381) ascribe this to potential skill degradation and loss of SA which in turn can be attributed to complacency and vigilance problems, passive instead of active information processing, and an alteration of the feedback provided by the system. In the following, sources from mainly the automotive domain and aviation are reviewed to analyse the challenges.

Skill Degradation

Skill degradation refers to the temporally increasing inability of human operators to perform the manual control of a task, e. g. the driving task (see Figure 2.8). Factors that are generally considered to affect the skill of operators are, for example, age, drugs, experience, but also automation (Trösterer, Gärtner et al. 2016).

Based on discussions with leading individuals involved in the administration of pilot training, Wiener and Curry (1980, pp. 1000–1001) noted a 'perceptible skill loss' as a consequence of the extensive use of automation systems in aircraft. Affirming this, the Federal Aviation Administration (FAA) more recently issued a safety alert that recommends pilots to take over manual control more often in order to prevent a 'degradation of the pilot's ability to quickly recover the aircraft from an undesired state (FAA 2013)'.

Casner, Geven et al. (2014, pp. 1514–1515) investigated the effect of automation on manual flying skills in a study with experienced airline pilots. While instrument scanning and manual control skills showed to be largely faultless, problems were observed in regard to cognitive abilities required for manual control. As such, the maintenance of awareness, e. g. regarding the position of the plane relative to the planned route without using a map display, was an aspect pilots struggled with. In an earlier study, Mengelkoch, Adams and Gainer (1971, p. 397) examined the retention of manual flying skills after a period of four months following the initial training. Supporting Casner, Geven et al. (2014), the results point towards a similar difference between hand-eye skills such as instrument scanning and tasks that required cognitive skills, such as recalling procedures. The general effect of deskilling due to automation has also been observed in other domains, e. g. engineering and manufacturing (Cooley 1987).

Regarding on-road vehicles, Trösterer, Gärtner et al. (2016) used an online survey that investigated the perceived driving skills of drivers that have either stopped driving, drive regularly, or have recently started to drive again. They measured the self-assessed perceptual-motor skills, i. e. regarding the basic vehicle motion control (operational functions, see Figure 2.8), and safety skills, i. e. the capability of driving safely (tactical functions), using the Driver Skill Inventory (DSI), a specialised instrument for assessing driving skills using surveys. Their findings suggest that, once acquired, perceptual-motor skills are retained during long periods of time without driving activity, similar to riding a bicycle. These findings are in line with the described retention of manual control skills in aircraft. The results of the online survey, however, rely on the drivers' ability to assess themselves as well as their honesty. Past research suggests that individuals perform badly at self-assessment (Dunning et al. 2003), with drivers mainly showing a positive self bias that leads them to perceive themselves as more capable than the average (McKenna, Stanier and Lewis 1991, pp. 49–50).

Currently, there are no alternative publications known to the author that analyse the deterioration of manual driving skills over time in regard to automated driving. The presented research can be seen as an indication that low-level skills, i. e. perceptual-motor skills that are needed to fulfil the operational functions of the driving task, are quite robust over time. Contrasting this, high-level, i. e. cognitive, more tactical skills, tend to deteriorate without practice. Especially in combination with the increasing complexity of automated systems, the loss of cognitive skills increases the need for supporting SA.

Loss of Situation Awareness

The loss of SA is considered to be the most significant aspect of the OOTL performance problem (Endsley and Kiris 1995). It leads to an increasing time necessary for the reorientation of the operator to relevant system parameters that enable problem diagnosis and manual control (Endsley and Kiris 1995, p. 382). The phrase *knowing what is going on around you* is often used to describe the concept of SA at the most basic level (Jones 2014, p. 98). A large number of models that characterise SA have been proposed. In an extensive review, Stanton, Salmon, Walker, Salas et al. (2017) analysed existing models of SA and placed them into three categories, each suited for specific HF problem areas: individual SA, team SA, and system SA. Table A.1 summarises their findings.

A major difference can be observed between individual and team SA. While individual SA only accounts for one human operator within the system, team SA additionally considers the information exchange between human team members, for example pilot and co-pilot in an aircraft (Salas et al. 1995). The latter concept is not relevant in the context of this research as only one human operator is present in a commercial on-road vehicle. System SA, on the other hand, considers the SA to be distributed within the whole sociotechnical system, e. g. between aircrew, aircraft sensors, and autopilot.

It has to be pointed out, however, that the described division of SA into three different models is controversial. Endsley (2015), for instance, argues that inanimate objects such as sensors do not have SA but are merely sources of information for the human operator. This view highly contradicts the model of system SA. Endsley (2015, p. 25) further argues that it is not sufficient that the awareness is *somewhere* in the system. While Stanton, Salmon, Walker and Jenkins (2010) state, for instance, that pilots are not required to remember the speed settings that underlie certain presets, Endsley (2015, p. 25) argues that exactly this is required for safe performance.

As of today, the SA concept proposed by Endsley (1995b) remains the most widely acknowledged within HF research (Golightly et al. 2010), as indicated by citation numbers (see Table A.1), and has been successfully applied to improve the design of various human-machine systems (Endsley and Jones 2012). Further, the model is particularly suited for this project as it clearly states the significance of well-designed interfaces and effective information presentation for SA (Endsley 2015, p. 16). Consequently, this thesis employs the model proposed by Endsley (1995b) which defines SA as 'the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future (Endsley 1988b, p. 792, 1988a, p. 97)'. As such, SA consists of three ascending, but not linear, hierarchical phases (Endsley 1995b, pp. 36–38, 2015, p. 8):

1. *Perception of the elements in the environment:* For the first level of SA, human operators need 'to perceive the status, attributes, and dynamics of relevant objects in the environment (Endsley 1995b, p. 36)'. A driver of an automobile, for instance, has to be aware of the status and movement of surrounding vehicles, potential obstacles, and the own vehicle as well as of cues in the environment, such as road signs.
2. *Comprehension of the situation:* Level 2 SA is based upon the first level and requires an understanding of the significance of the perceived elements and their combinations in regard to the objectives of the user. A vehicle driver on a motorway, for example, might have perceived the acceleration of a vehicle in the adjacent fast lane immediately before a motorway exit and can conclude the intentions of the other road user.
3. *Projection of future scenarios:* Level 3 SA refers to the ability of projecting the status of the elements in the (near) future based on the knowledge acquired in the previous two levels. Continuing with the example above, the vehicle driver can predict that the other road user might change lanes, decelerate abruptly and exit from the motorway.

The acquisition of SA forms the basis for decision-making and, subsequently, the performance of actions (see Figure A.3). While the acquisition of ideal SA supports both decision-making and performance, it does not guarantee it. Other factors, such as experience or training, influence the decision-making process and individuals might be unable to perform in an intended way. Besides, SA is limited to aspects of the dynamic environment and does not comprise static knowledge such as checklists (Endsley 1995b, pp. 36–40).

As indicated by the presented definition, SA is a highly temporal construct. SA can be acquired over time rather than instantaneously and encompasses knowledge about elements in the past as well as the future (Endsley 1995b, p. 36). Thus it can be argued that there is a certain *buffer* that is being reduced and filled relative to the operator's attention towards the task to be performed (Kircher and Ahlström 2009). In the following, the specific HF problems that lead to a loss of SA or impede its acquisition are reviewed.

Complacency, Automation Bias, and Vigilance Problems Operators of highly automated systems are often demanded to monitor system performance over an extended period of time. This prolonged monitoring task requires sustained attention (see Subsection 2.4.3) in order to perceive relevant events such as system failures. The term *vigilance* refers to 'the action or state of keeping careful watch for possible danger or difficulties (Oxford Dictionaries 2016b)', thus it refers to the ability to retain sustained attention on the task at hand. Operators with sufficient vigilance levels – and therefore likely sufficient SA – will presumably notice potential system malfunctions or abnormalities and can subsequently react to them. However, the prolonged monitoring and the sustained allocation of attention to the task leads to a decrease in vigilance levels, referred to as vigilance decrement (Parasuraman 1986). Thus, over time, the ability of operators to detect relevant events decreases.

In a recent driving simulator study, Greenlee, DeLucia and Newton (2018) found supporting evidence. A total of 22 individuals were seated in a simulated vehicle with activated driving automation and were tasked with the detection of hazardous roadway events. In concordance with the described vigilance decrement, the monitoring task during automated driving led to decreases in performance over time.

The vigilant decrement is thereby further aggravated by users' overreliance in and misuse of automation (Endsley 2017). In practice, cases of automation misuse and overreliance manifest themselves in the user's 'uncritical reliance on the proper functioning of an automated system without recognizing its limitations and the possibilities of automation failures (Manzey, Reichenbach and Onnasch 2012; Parasuraman and Riley 1997)'.

Complacency and *automation bias* are indications of automation misuse. The former can be characterised as 'an active reallocation of attention away from the automation to other tasks [. . .] under conditions of multiple-task load, when manual tasks compete with the automated task for the operator's attention (Parasuraman and Manzey 2010, p. 384)'. Consequently, complacency can be considered a predictor for visual attention (Bailey and Scerbo 2007; Metzger and Parasuraman 2005; Parasuraman and Manzey 2010) and, therefore, impacts failure detection. Insufficient vigilance and a monitoring frequency below an optimal standard are the consequence of complacency, resulting in inappropriate levels of SA (Geiselman, Johnson and Buck 2013, p. 25). Research findings suggest that particularly highly reliable automation systems that fail only on very rare occasions – like automated driving systems – promote complacency, while increasing experience and practice showed no significant effect (Moray and Inagaki 2000; Parasuraman and Manzey 2010; Waard et al. 1999; Bailey and Scerbo 2007; Molloy and Parasuraman 1996). In addition to automation reliability (Bagheri and Jamieson 2004; Wickens, Sebok et al. 2015), the fluctuation of that reliability affects complacency (May, Molloy and Parasuraman 1993). Automation bias, on the other hand, refers to the tendency of human operators to place greater faith in the automation than in other information sources. Consequently, users do not respond to critical situations if not alerted (error of omission) or follow an erratic advice (commission error) (Parasuraman and Manzey 2010).

Several system-related, individual, and context-dependent factors influence complacency and automation bias. Particularly an overtrust in the capabilities of the system is considered to be a key reason for both (Parasuraman and Manzey 2010; Singh, Molloy and Parasuraman 1993, pp. 118–120). Thus, trust indirectly affects the attention management, vigilance, and SA of the human operator. This is further supported by thematic video analysis results of an on-road driving study involving Tesla Model S drivers. Instead of monitoring the system behaviour in accordance with the actual reliability of the vehicle, drivers allocated their attention away from the driving task, showing behaviour characteristic for complacency issues (Banks et al. 2018). The case of a Tesla Model S user found sleeping in his moving vehicle while the automated driving system was presumably activated provides a real-world example of automation misuse (Davies 2018). A potentially interrelated concept is that of *risk homoestasis* which leads operators to aim for a preferred risk level. As the automation is perceived as increasingly capable due to longer periods of flawless performance, the operators could tend to rely more on the automation system and, as a consequence, monitor it less (Cummings and Ryan 2013, p. 5).

However, even if operators do not exhibit complacent behaviour and maintain an optimal attention allocation strategy, relevant events such as system failures may be missed (Moray and Inagaki 2000). Allocating attention away from the system and placing attentional resources on competing tasks can be an optimal strategy for highly reliable automated systems that fail only very rarely. This strategy, albeit optimal, places attention away from the automation for the majority of time. As a consequence, it is highly unlikely that users monitor the system exactly at the moment in which it fails.

Subsection 2.4.3 has introduced the concepts of top-down and bottom-up attention allocation. The attention allocation strategy affected by automation reliability can be classed as a top-down mechanism, whereby the operator's attention management depends on the perceived capabilities of the automated system. In order to support users and attract attention in critical events, exogenous stimuli are necessary (bottom-up attention allocation). In this context, the detrimental side effects of alarm-like stimuli have to be accounted for, e. g. the *cry wolf effect* (Yang et al. 2017).

Passive Information Processing The passive monitoring role of human operators in automation systems leads to a lower degree of cognitive engagement which makes it more difficult for (even attentive) users to understand the current situation and gain SA (Endsley 2017, p. 8; Endsley and Kiris 1995). Wickens and Kessel (1979), for example, observed that human operators using an autopilot are slower to detect problems compared to those that are actively processing information in manual control.

Alteration of Feedback In vehicles equipped with automated driving systems, a different kind of feedback is required than in vehicles without driving automation systems. In the latter, the human operator determines – and is therefore aware of – the current and future movement of the vehicle. In the former, however, this information should be communicated to the user. Endsley (2017, p. 7) names several examples from aviation in which the system failed to provide necessary feedback, leading to accidents. Some information is deliberately hidden from the operators in order to keep the interface understandable, other deletions are unintentional. Without an appropriate level of transparency, operators cannot achieve SA about the status of the system and its intentions (Endsley 2017, p. 7). The transparency of automated systems will be reviewed in Section 2.8.

2.6.2.3 Trust and Transparency as Key Challenges

This section has established the OOTL performance problem as a major challenge when designing for conditional driving automation. Particularly a lack of SA impedes the safe manual driving performance following takeovers. As a consequence of the reasons for insufficient SA, the acquisition and maintenance of SA can be supported by

- preventing the user's overtrust in the capabilities of the system and
- by providing feedback about the system state and its intentions.

These fundamental challenges provide opportunities for counteracting the OOTL performance problem and thereby promote a safer driving performance following takeovers.

2.7 Designing for Appropriate Trust in Automation

As outlined in the previous section, trust is a key concept for counteracting the OOTL performance problem and thereby achieving a safer driving performance. This section provides an introduction to the concept of trust and analyses how appropriate trust can be achieved.

2.7.1 Definition and Concept

Trust is a social psychological concept that can be defined as 'the attitude that an agent will help achieve an individual's goal in a situation characterized by uncertainty and vulnerability (Lee and See 2004)'. This definition deliberately uses the broad term *agent* to indicate that trust can not only be placed in humans but in any entity that interacts with the environment in an active manner. In fact, similar processes are involved in trust between humans and trust between humans and human-like machines (Visser, Monfort et al. 2017), rendering efforts for maintaining human-machine relationships such as trust repair mechanisms necessary (Visser, Pak and Shaw 2018). Further, uncertainty and vulnerability are important aspects as trust helps users to reduce the cognitive effort in complex situations and guides user reliance on automation (Lee and See 2004).

The increasing sophistication of automated systems makes a complete understanding of the automation and its underlying processes unrealistic, rendering trust an important mediator in the interaction process between human and machine (Lee and See 2004). In several ways, trust was shown to guide the reliance on and use of automation (Dzindolet et al. 2003). Particularly the monitoring behaviour of operators using automated systems is affected by trust. Both Muir and Moray (1996) and more recently Hergeth et al. (2016) identified an inverse relationship between operator trust and the monitoring activity of users regarding task-related AOIs. Payre, Cestac and Delhomme (2016) examined the relationship between operator trust and reaction times following takeovers and found that higher trust leads to longer reaction times. As such, trust is key to avoid misuse (induced by overtrust) and disuse (induced by distrust) (see Figure 2.12) (Parasuraman and Riley 1997). This is in line with the findings described by Parasuraman and Manzey (2010) who recognised overtrust as a key culprit for complacency and automation bias (see Subsection 2.6.2.2). The tendency of users to overtrust automated driving systems was highlighted by a recent photograph of a sleeping driver in a moving Tesla Model X equipped with a level 2 driving automation system (Cuthbertson 2019). Nonetheless, appropriate trust does not guarantee appropriate reliance but rather guides it (Chavaillaz, Wastell and Sauer 2016). Other factors such as workload, self-confidence, and time constraints affect the decision to rely on a system in addition to trust (Lee and See 2004).

2.7.1.1 Appropriate Trust

Appropriate trust can potentially counteract the outlined problems and promote sufficient monitoring behaviour. Trust is given the attribute *appropriate* when it matches the true capabilities of the agent. In an extensive literature review regarding trust in automation,

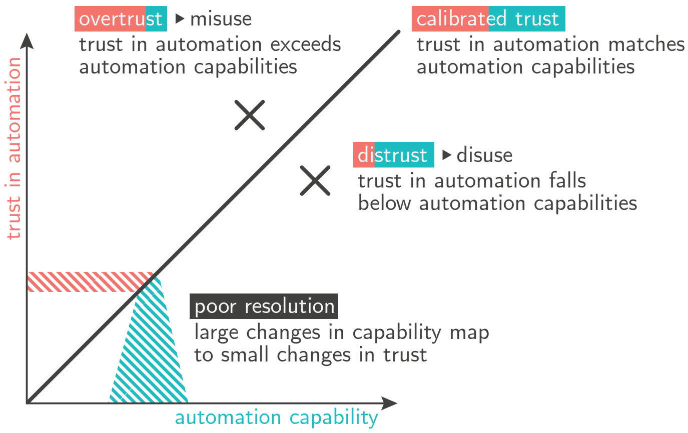


Figure 2.12: Schematised description of calibrated trust and trust resolution (adapted from Lee and See 2004)

Lee and See (2004) identified three essential aspects for achieving appropriate trust (see Figure 2.12):

1. calibration: agreement between the user's trust in automation and the capabilities of the automation (Lee and Moray 1994; Muir 1987)
2. resolution: degree to which the user's trust judgement differentiates different capability levels (Cohen, Parasuraman and Freeman 1997); with high resolution, for instance, changes in capability are reflected in equivalent changes of trust
3. specificity: degree of differentiation between different components or aspects of the trustee
   a) functional specificity: degree of differentiation between functions
   b) temporal specificity: sensitivity to changes

A combination of calibrated trust as well as high resolution and specificity can alleviate both disuse and misuse. When interacting with an automated system, the HMI can support users with the acquisition of appropriate trust. Lee and See (2004) define the information that the system can provide along two dimensions:

1. abstraction: information regarding the performance, process, and purpose of the automation (Lee and Moray 1992)
2. detail: describes the entity that is to be trusted, ranging from the overall system to its single functions and modes

Combined, abstraction and detail form the basis of trust. *Performance* refers to the automation's historic and current perceivable operation, answering the question *what* the system does. Supporting information are, for example, historic failure rates. *Process* is the appropriateness of the machine's internal operations for achieving the trustor's goals and answers *how* the system operates. In this regard, users could be supported by explanations about the automation's decision-making process. *Purpose* is an interpretation of the designer's intended use of the system, referring to *why* the system has been developed (Lee and Moray 1992).

Regarding the dimension *detail*, the system-wide theory argues that operators merge the trust of independent functions of the same system. Thus, a higher detail may not lead to more appropriate trust (see Subsection 2.7.1.2) (Keller and Rice 2010).

2.7.1.2 Trust in Complex Systems

Complex systems such as aircraft cockpits or, increasingly so, automated driving systems may support users with multiple aids to perform certain tasks. These aids may thereby not have concordant reliability levels. Rather, the aids likely differ in terms of reliability. For instance, a lane departure warning system may have a low reliability if the lane markings are faded. The subsystem responsible for longitudinal guidance and control (ACC), on the other hand, is unaffected by this and might simultaneously have a high reliability. This poses the question if users trust each aid according to its reliability level (component-specific trust) or if they merge their trust across each aid to form a trust level for the overall system (system-wide trust). This question relates to the level of detail that the trust is based on (see Subsection 2.7.1.1).

Thinking of inter-human interaction, one may jump to the conclusion that trust is dependent on the task at hand. A co-worker might be very competent at solving differential equations but may not be your first choice when looking for someone to proofread a document. In everyday interaction, we can effortlessly distinguish between the function-specific capabilities of humans around us – but does this remain true for automated agents?

According to the component-specific trust theory, the reliability of one particular aid does not affect the users' trust of the others. Applying this theory, users can distinguish between the reliability levels of automated aids without contagion effects. When interacting with two aids in a complex system, for instance, users would use each aid according to the perceived reliability level (Keller and Rice 2010).

System-wide trust theory, on the other hand, predicts that users perceive a complex system as one unit rather than a composition of several individual components. Thus, when interacting with two aids differing in reliability, operators will use them as if they had the same reliability level.

Figure 2.13: Screenshot of the tracking task employed in the experiments presented by Keller and Rice (2010)

Keller and Rice (2010) investigated the presented theories in a study involving 36 participants who had to monitor two horizontally aligned gauges while performing a tracking task (see Figure 2.13). Specifically, participants were required to notice system failures that manifested themselves as values above or below an acceptable range. The detection of system failures was supported by an automated aid for both gauges. The monitoring of the right-hand gauge was supported by a perfect automated aid with a reliability of 100 %. The reliability of the other aid was varied between-participants (70 %, 85 %, 100 %). For conditions in which a perfect aid was paired with an imperfect aid, monitoring performance metrics did not significantly vary between gauges. Employing component-specific trust theory, the monitoring performance regarding the gauge supported by the perfect aid should have been superior to that of the other gauge. Contradicting this, the less reliable aid seemed to have pulled down the trust level for the highly reliable aid. Thus, users merged the perceived reliability across aids and did not consider each aid independently, favouring system-wide trust theory.

However, the merging of trust may have been supported by similar visual layouts and the proximity of the gauges to each other as well as the fact that the automated aid used the same modality (auditory) and also the same voice for both gauges. Further, only two functions were supported by automated aids. Operators could assume that the failure of one aid affects the whole system as it makes up 50 % of it. Additionally, the study does not provide insights if the users' reliance on the aids changed over time. Moreover, an indication of the reliability of each aid may have supported the user in establishing appropriate trust levels for each aid. It can therefore be argued that the focus of trust 'lies on a continuum (Keller and Rice 2010, p. 125)', depending on the context of use, rather than being either component-specific or system-wide.

In a follow-up study, Rice and Geels (2010) increased the number of aids from 2 to 4, removed the tracking task to reduce the effects of a second activity, and introduced the variables *reliability information* and *feedback*. The gauges were arranged similar to the layout presented by Keller and Rice (2010) and participants were again asked to detect whether the gauge displayed a value within the safe range. A total of 99 participants took part in the experiment, whereby the amount of information available to the participants was varied between-subjects. In the first condition, participants were told that the reliability of the gauges was unknown and received no feedback of their performance. In the second condition, participants were notified about their decision accuracy after each trail. In the third condition they were previously told the reliability of each gauge. For all three conditions, the gauge supported by the imperfect aid (70 %) was placed on the leftmost position, with all other gauges being 100 % reliable. Finally, a control condition was implemented that presented participants with four aids that were all 100 % reliable.

The authors used dependence, i. e. the agreement rates between automated aids and participants, as an indicator for the employment of a system-wide trust strategy. When the aid correctly determined that the system was *safe*, the agreement rates did not significantly differ between each gauge, i. e. there was no main effect of the variable *reliability*. As agreement, and thereby dependence, should vary between gauges, the authors concluded that participants employed a system-wide strategy. However, a lower agreement score would mean that participants incorrectly rejected the aid's recommendation. Thus, participants that were comparing the aid's decision with their own observation would (correctly) conclude that the aid is correct. Consequently, the agreement rates for situations in which the gauge incorrectly indicated that the gauge was *safe* are of higher interest. In this situation, the agreement rate was significantly lower, particularly when the reliability information was available. While this may indicate lower trust in this particular aid, it might also indicate that participants were routinely comparing their own observations with the recommendations of the automated aid – or might not have considered the aid's decision at all. The study lacks eye tracking data to reject or confirm this. Further, the study also indicated that providing participants with reliability information increases both accuracy and correct dependence while decreasing incorrect dependence. Thus, the provision of reliability information may support users in employing a component-specific trust theory.

Under consideration of the compliance-reliance paradigm (Meyer 2001, 2004; Chancey et al. 2017), Geels-Blair, Rice and Schwark (2013) conducted an additional study to investigate the impact of false alarms (FAs) and misses on the employment of the system-wide theory. The compliance-reliance paradigm refers to the behaviour of an operator when interacting with an automated aid. *Compliance* is the operators' response to a signal issued by the aid.

Reliance is the absence of a response when no signal is issued, i. e. for normal operation. The former is affected by FAs, the latter by misses.

In accordance with Rice and Geels (2010), participants were required to monitor gauges, with the difference that 8 instead of 4 gauges were used. Again, only the leftmost gauge varied in reliability while all remaining gauges were 100 % reliable. The study implemented three reliability conditions: 100 % reliable, 70 % reliable with FAs, 70 % reliable with misses. Prior to each experiment, participants were informed about the reliability level of each gauge and, if applicable, the failure type.

As predicted, the FA prone aid resulted in significantly lower agreement rates for compliance, while the aid prone to misses showed a decrease in agreement rates for reliance. Compared to the results of the previously presented studies (Keller and Rice 2010; Rice and Geels 2010), there was no strict use of the system-wide trust theory apparent. Nonetheless, the fact there was some contagion despite the presence of seven perfectly reliable aids shows that lower reliability levels of one aid affect the use of other aids, thereby implying the employment of a system-wide trust theory. The results further confirm the theory proposed by Keller and Rice (2010) stating that the focus of trust lies on a continuum. Thus, trust is not exactly merged across all aids but affected nonetheless. Further, knowledge of the component-specific reliability was shown to facilitate component-specific trust.

2.7.1.3 Trust Formation

Hoff and Bashir (2015) extended the review of Lee and See (2004) and established a trust model consisting of three interdependent trust levels: dispositional trust, situational trust, and learned trust – each of which are affected by different factors.

Dispositional trust is influenced by a person's propensity to trust as affected by age, culture, gender, and personality traits. Regarding *situational trust*, it can be differentiated between external and internal variability. The former refers to the system type, its complexity, task difficulty, workload, perceived risks, perceived benefits, the organisational setting, and the framing of the task. The latter includes self-confidence, subject matter expertise, mood, and attentional capacity (Hoff and Bashir 2015).

Hoff and Bashir (2015) define *learned trust* as a representation of the users' evaluations based on prior experience and current interaction. Further, it is distinguished between initial learned trust and dynamic learned trust, with the former existing *before* an interaction and the latter developing *during* an interaction. Initial learned trust is mainly affected by attitudes and expectations regarding the system, its reputation, prior experience, and the understanding of the system. Dynamic learned trust is mainly influenced by the users' perceptions of system performance which are in turn dependent on its design features.

As such, anthropomorphism, ease of use, communication style, transparency, and level of control have shown to be significant design features that impact trust (Hoff and Bashir 2015, p. 425). Particularly dynamic feedback regarding the automation's reliability as well as situational factors that influence automation performance were recognised as key aspects.

In a meta-analysis of factors affecting trust in human-robot interaction, Hancock et al. (2011) identified the performance of the automation as the main factor that influences trust. Contrasting this, Balfe, Sharples and Wilson (2018) found that an understanding of automation and system transparency are most relevant for trust in real-world settings, highlighting potential differences to lab-based experiments. In another meta-analysis regarding factors influencing the development of trust in automation, Schaefer et al. (2016) confirmed the importance of transparency for trust development and simultaneously urged for more research regarding this topic. In an additional meta-analysis, Wickens and Dixon (2007) identified that the presence of competing tasks leads to an attention allocation away from the automation, which is further exacerbated with higher levels of automation (Carsten et al. 2012) and affected by the type of concurrent tasks (Diekfuss, Ward and Raisbeck 2017). Moreover, the attention allocation strategy is affected by operator workload (Young and Stanton 2002).

This emphasises that a comprehensive view is required to investigate how human-machine interfaces can impact trust in automation and thereby SA.

2.7.2 Integral Model

Figure 2.14 illustrates the formation of trust and its significance for the acquisition of SA under consideration of related concepts such as workload. The schema is separated between the user perspective and the automation. Path 1 shows that certain HMI aspects such as system transparency have a major influence on the dynamic learned trust, as outlined in Subsection 2.7.1.3 (Schaefer et al. 2016; Hoff and Bashir 2015; Balfe, Sharples and Wilson 2018). The second path illustrates the determination if the trust of the user is appropriate (see Subsection 2.7.1.1) (Lee and See 2004). In case the trust falls below the capability of the automation, the user will likely disuse the system (see Subsection 2.7.1) (Parasuraman and Riley 1997; Lee and See 2004). A trust that exceeds the capabilities of the system induces both complacency and an attentional bias (see Subsection 2.6.2.2) (Parasuraman and Manzey 2010). This leads to insufficient monitoring behaviour (Hergeth et al. 2016) and thereby a lack of SA (Endsley 2017). In the event that the trust can be considered appropriate, the likelihood of sufficient SA increases.

While the schema indicates that individual differences impact the acquisition of SA, the HMI can be designed to support the appropriate trust formation independent from

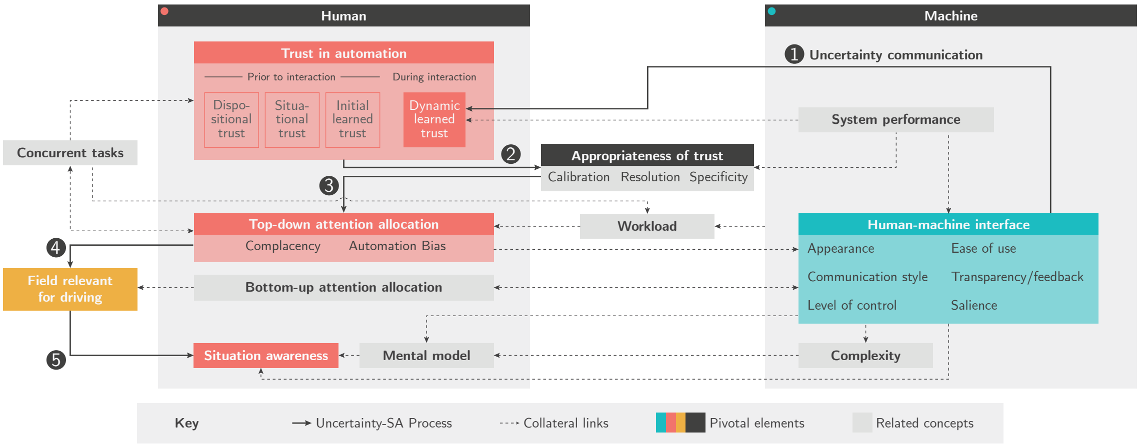


Figure 2.14: Schematised model illustrating the significance of appropriate trust (own representation based on Endsley 2017; Schaefer et al. 2016; Hoff and Bashir 2015; Parasuraman and Manzey 2010; Lee and See 2004; Parasuraman and Riley 1997)

individual differences. The meta-analysis conducted by Hoff and Bashir (2015) resulted in HMI aspects that have been empirically proven to influence trust, whereby the factor *transparency* was identified most frequently (18 out of a total of 35 publications).

2.8 System Transparency

System transparency was shown to have a major impact on the formation of dynamic learned trust and is therefore a key influencer of achieving sufficient SA. This section defines system transparency and reviews recent approaches to achieve system transparency.

2.8.1 Definition and Benefits

Several definitions of transparency in the context of automation exist. According to Endsley (2017), the transparency of automated systems 'refers to the understandability and predictability of their actions.' Elaborating on this definition, Chen, Procci et al. (2014) define transparency along the three-level SA model (see Subsection 2.6.2.2): Transparency is 'the descriptive quality of an interface pertaining to its abilities to afford an operator's comprehension about an intelligent agent's intent, performance, future plans, and reasoning process'. As such, the concept of system transparency is consistent with the previously introduced basis of trust (see Subsection 2.7.1.1).

The definitions highlight that increasing transparency does not imply to render all information pertaining to an inanimate agent accessible to the user. Rather, increases in transparency aim at improving users' understanding of a system, i. e. their mental model, without overburdening them. Currently available vehicles rely on user manuals to convey this information – which was found to be insufficient for vehicles equipped with automated driving systems (Boelhouwer et al. 2019). Instead, systems should provide real-time information to afford users a situation-dependent calibration of their expectations and to foster a more appropriate mental model (Seppelt and Lee 2007).

Moreover, transparency may mitigate the cry wolf effect (Yang et al. 2017), i. e. the tendency of operators to neglect alarms as the consequence of a high false alarm rate, particularly in high risk systems were the alarm threshold is set deliberately low to prevent critical information not being noticed (Breznitz 1984).

2.8.2 Uncertainty Communication

The performance of automated systems was shown to have a major impact on the dynamic learned trust of operators (Hancock et al. 2011; Schaefer et al. 2016). Providing information that allows users to assess the current capability can enhance the users' judgement about the system's performance. In this context, several publications have explored the communication of system uncertainties to render the system's current state somewhat transparent (Chen, Lakhmani et al. 2018). In reference to Section 2.2, uncertainty is regarded as a system parameter that characterises the confidence associated with automation processes and outputs, specifically concerning the DDT. As such, the term is closely related with concepts such as automation reliability and capability. In the following, existing experimental work pertaining to uncertainty communication within the automotive and other domains is discussed.

2.8.2.1 Uncertainty Communication in Automated Driving

Several research attempts have focused on conveying information about the automation's uncertainty to support the trust calibration of users in the context of automated driving. Beller, Heesen and Vollrath (2013) investigated the impact of displaying a schematised uncertain face in the instrument cluster in unclear situations (see Figure 2.15a). The results of a driving simulator study indicate that the communication of system uncertainties has a positive effect on metrics relating to driving safety, in particular time to collision (TTC). Besides benefits to performance metrics, the provision of uncertainty information led participants to pursue a more appropriate attention allocation strategy relative to the control group. In accordance with the integral model depicted in Figure 2.14, attention was directed towards the FRD prior to critical situations, resulting in improved SA and supporting driving safety.

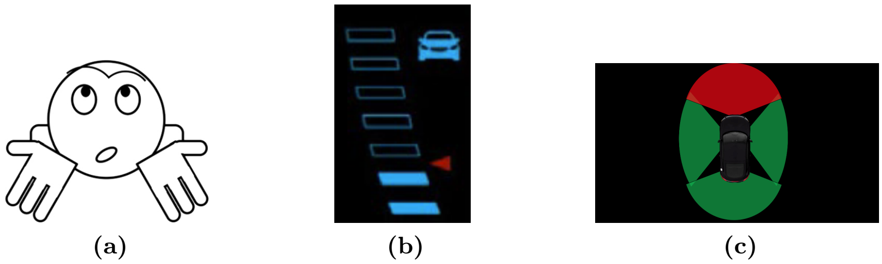


Figure 2.15: Uncertainty visualisations in the autmotive domain (Beller, Heesen and Vollrath 2013; Helldin et al. 2013; Large, Burnett, Salanitri et al. 2019)

Supplementing this work, Helldin et al. (2013) explored the communication of seven different uncertainty levels using bars presented in a digital instrument cluster display, each bar representing one level (see Figure 2.15b). The results show that users who were presented with the uncertainty information could afford to allocate their attention away from the FRD and perform NDRTs for a longer time than the control group. Nonetheless, the participants of the experimental group were able to take over the DDT faster than those of the control group.

In a more recent approach, Large, Burnett, Salanitri et al. (2019) conducted a longitudinal driving simulator study that replicated 30-minute commute journeys on five consecutive days. Participants interacted with an SAE level 3 automated driving system and were given the option to engage in an NDRT of their choice during activated driving automation. On day four of the experiment, participants had to perform an emergency takeover. Using a between-subjects design, one group of participants was provided with sensor status information on a display located in the centre console. Colour, specifically a change from green via amber to red, was used to convey an increase in severity for each vehicle sensor (see Figure 2.15c). The provision of status information resulted in significantly reduced time until readiness to drive was achieved following the TOR. In the context of this study, readiness to drive was defined as the time taken until both a glance to the FRD was performed and at least one hand was on the steering wheel.

Whereas the previously outlined publications focused on the implications of uncertainty communication on operator behaviour and driving metrics, Noah, Gable, Chen et al. (2017) explored the use of different visual display types for conveying the reliability of an automated lane keeping system. Specifically, human-centric and system-centric displays using qualitative, quantitative, and representational system reliability visualisations were assessed in terms of their ordinal characteristics using a forced choice sorting task. While there was a similar matching accuracy for qualitative, quantitative, and representational display types, participants could more readily match system-centric displays with the intended reliability levels. Consequently, HMIs should convey the state of the system, for instance *moderate uncertainties*, rather than required user behaviour.

Research Gaps

The presented publications affirm the outlined benefits of communicating uncertainties (see Figure 2.14). A limitation of previous work, however, is that users are required to move their focus towards the instrument cluster in order to gain knowledge about the system's current uncertainty. Already, this has shown benefits regarding the practicability of NDRTs (Helldin et al. 2013), but solutions that do not require the driver to glance to the instrument cluster will likely improve this further. The additional monitoring demands also increase information access effort (see Figure 2.7), likely leading to workload increments. In fact, a more comprehensive investigation of uncertainty communication is needed that takes all key parameters depicted in Figure 2.14 model into account. As such, the temporal course of operator trust relative to uncertainty levels is of particular interest to assess trust calibration (see Subsection 2.7.1).

Existing research also lacks an investigation of different levels of functional specificity. In the experiments conducted by Beller, Heesen and Vollrath (2013) and Helldin et al. (2013), the system uncertainties were merged into a system-wide metric. As a higher functional specificity was suggested to support appropriate trust in automation (see Subsection 2.7.1), more research efforts are required in the context of driving.

Further, the studies did not explore and evaluate different modalities of conveying uncertainties. In reference to the multiple resource model (see Figure 2.6), the uncertainty information should address a sensory channel that is not preoccupied with other tasks. Consequently, particularly channels that are likely not addressed by DRTs or NDRTs should be used to convey information, for instance using peripheral or vibro-tactile cues (see Subsection 2.5.3.3).

To supplement the presented findings relating to uncertainty communication in the context of automated driving, publications originating from other domains will be evaluated, with a particular emphasis on aviation (see Subsection 2.6.1).

2.8.2.2 Uncertainty Communication in Other Domains

In contrast to automated driving systems, automated systems in other domains are often implemented as decision aids that do not perform actions on behalf of the user. In the following, publications mainly from the military domain, aviation, and data visualisation are reviewed to gain insights that are not currently available in the automotive sector.

Within the military domain, automated aids are frequently used to support the identification of hostile objects or soldiers. In this context, Dzindolet et al. (2003) conducted three interrelated studies to investigate the role of trust in automation reliance. In each experiment, participants were presented with a series of pictures showing potential combat scenes. For each picture, participants had to determine if a camouflaged soldier is present. An automated aid supported participants by providing decision recommendations. Following each picture, participants were asked if they comply with the recommendation of the automated aid. In two of the studies, participants were provided with bar diagrams displaying the amount of mistakes made by the system and the user, thus increasing the transparency regarding the system's performance. The results indicate that the provision of continuous feedback regarding the performance of the aid can support trust calibration as well as appropriate reliance. Besides, the third study showed that the provision of a rationale for errors increases both trust and reliance independent of the actual reliability of the aid, which is in line with previously described findings (see Subsection 2.5.3) (Koo et al. 2015).

Wang, Jamieson and Hollands (2009) conducted a complementary study regarding a combat identification (CID) aid. A CID system aids an infantry soldier with the identification of other soldiers by displaying a blue light for *friendly* and a red light for *unknown* encounters. While the blue light reliably indicates a friendly soldier, the red light could mean hostile, friendly, or civilian. In their experiment, Wang, Jamieson and Hollands (2009) examined the impact of verbally disclosing a constant percental value as a representation of aid reliability on trust and reliance. Their results indicate that both dependent variables were positively affected through an improved awareness of aid reliability.

Neyedli, Hollands and Jamieson (2011) expanded on this research. While Wang, Jamieson and Hollands (2009) verbally provided the reliability once for each block of trials, Neyedli, Hollands and Jamieson (2011) presented the reliability information in real-time, arguing that the reliability is a volatile variable that changes with distances or environmental conditions such as humidity. For this application, four prototype visualisations varying in display type and proximity were developed and evaluated (see Figures 2.16a to 2.16d):

- variations in display type, i. e. the visualisation technique:
    1. pie chart; proven as a relatively accurate method to depict proportions ($\pm 2 - 3\,\%$ (Hollands and Spence 1998))
    2. random mesh display consisting of an arbitrary arrangement of small squares in a grid, with the proportion of filled squares proportional to the reliability; potentially more effective method to convey information in a time-critical situation (as suggested by Chong and Treisman (2003, 2005) and Ariely (2001))

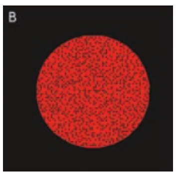
- variations in information proximity, i. e. the relative location of presented information:
  1. integrated depiction of identification info (blue/red) and reliability information (pie chart/mesh), known as spatial integration (Wickens and Carswell 1995)
  2. separated depiction of the information types; one circle coloured according to identification, another displaying the reliability in yellow (pie chart/mesh)

The four variations were investigated in a simulated combat environment. Quantitative results show that the random mesh display and the integrated display type led to the most appropriate reliance on automation and highest sensitivity to changes.

McGuirl and Sarter (2003, 2006) also investigated the impact of dynamically conveying confidence information on operator trust and performance, but their application was within aviation. Specifically, the authors investigated a decision aid based on neural networks to support the pilot in case of in-flight icing. They found that a line graph, or trend display, depicting the current and temporal development of system confidence improved the performance and trust calibration without adding excessive attentional demands. This is in line with the findings of Mercado et al. (2016) who also found that the communication of uncertainties does not in itself increase the workload of operators (see Subsection 2.8.1).

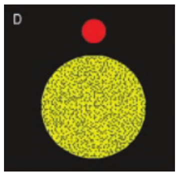
Finger and Bisantz (2002) examined the usability of degraded, i. e. pixelated, icons for communicating uncertainty by conducting two studies regarding CID aids. The results of the first experiment show that participants can distinguish between different pixelation levels and are capable of matching the icons to the associated uncertainty level. The second study aimed at assessing the differences in performance between the use of degraded icons, numeral probability values, and degraded icons with probability values. The performance of participants that were provided only with the degraded icons was superior to the other display types in some performance measures and comparable in others. The results suggest that degraded icons are a practicable means of conveying uncertainty, especially when the dynamic communication is required via an already cluttered display.

(a) pie chart, integrated



(b) mesh, integrated

(c) pie chart, separate



(d) mesh, separate

Figure 2.16: Variations of display type and proximity in CID (Neyedli, Hollands and Jamieson 2011)

Extending the findings of Finger and Bisantz (2002), Bisantz, Marsiglio and Munch (2005) explored the effect of using pixelated icons on the performance of participants and compared the results to the performance with numeric (e. g. 50 % uncertainty) as well as linguistic (e. g. *tossup*) expressions for uncertainty. Additionally, the resolution of the provided information was varied, i. e. either 5, 7 or 13 representations for differentiating between uncertainty levels. In three experiments, participants used a stock purchasing aid that provided information about the likelihood of profitability for each stock. The performance using the different display formats was similar, indicating that vague uncertainty representations like degraded icons are just as easily understood as precise representations like numeric expressions. This is in line with the findings of Finger and Bisantz (2002). Besides, the results indicate that the resolution of confidence information has an impact on performance, with higher resolution leading to improved results. An upper limit for resolution, however, was not defined. In a fourth experiment, Bisantz, Marsiglio and Munch (2005) prompted participants to generate fuzzy membership functions that show to what degree individuals associate particular representations or expressions with a probabilistic value. The generated functions showed high degrees of correlation between the maximum values of graphical/linguistic representations and the corresponding numerical values. Bisantz, Marsiglio and Munch (2005) attribute this to the hypothesis that people translate the vague representations, e. g. icons, to precise, numeric values.

The previously described publications focused on state uncertainty, i. e. uncertainty in regard to *what* something is (Bisantz, Marsiglio and Munch 2005). Several other studies investigated the communication of spatial uncertainty, i. e. *where* something is. Kirschenbaum and Arruda (1994) compared the suitability of graphically and linguistically communicating the uncertainty in the position of nearby submarines. The results favour the use of graphical representations (elipses) for spatial problems. In two complementary collision avoidance experiments within aviation, Andre and Cutler (1998) compared the use of numerical values, colours, and spatial information (ring size proportional to uncertainty). The results indicate similar conclusions as Bisantz, Marsiglio and Munch (2005), favouring graphical visualisations for spatial problems.

The largest body of research regarding the communication of uncertainty is available in the field of data visualisation. Data visualisation is concerned with the graphical display of data and receives inputs from several disciplines including computer science, psychology, and multimedia design (Aparicio and Costa 2014). In this sense, uncertainties 'include statistical variations or spread, errors and differences, minimum-maximum range values, and noisy or missing data (Pang, Wittenbrink and Lodha 1997, p. 371).' Uncertainty visualisation is of particular relevance for the subsidiary academic discipline of geographic information science (GIScience). GIScience is concerned with the study of geographic information systems (GISs), computational tools for analysing, storing, manipulating, and visualising spatial or geographic data on a map (GISGeography 2017; MacEachren, Robinson et al. 2005). A major challenge for GIScience is the integrated depiction of data and its uncertainty as well as the mapping of uncertainty components to data types. Thomson et al. (2005) proposed one of the most comprehensive typologies that maps seven uncertainty components, including accuracy, precision, and completeness, to the common data types in GIS – attributes, location, and time. This renders uncertainty visualisation within GIScience highly specific and requires various distinguishable symbols for depicting the individual uncertainties. Symbols, signs, and their meaning are the research subject of semiotics. Thereby is often discriminated between signifiers, i. e. the material form of signs, and the signified, i. e. the represented phenomenon or process (Chandler 2017; Roth 2017). Inspired by semiotics, Bertin (1967) introduced the concept of fundamental visual variables which can be manipulated to graphically convey information. The author proposed the following primitive signifiers:

- position: changes in the x, y, (z) location
- size: change in length, area, or repetition
- shape: infinite number of shapes
- value: changes from light to dark
- orientation: changes in alignment
- colour: changes in colour hue with a given value
- grain/texture: changes in pattern

Morrison (1974) (arrangement and saturation) and MacEachren (1992) (crispness, transparency, and resolution) added five further variables, the latter specifically for the purpose of depicting uncertainty (Halik 2012):

- arrangement: location of component elements
- saturation: desaturating colour with higher uncertainty
- crispness: contour crispness and fill clarity, out-of-focus metaphor
- transparency: overlay with increasing opacity for higher uncertainty, fog metaphor
- resolution: increasing pixelation with higher uncertainty

Figure 2.17 illustrates the visual variables. It must be noted that the implementation of the variable transparency is not unambiguous. Instead of using an additional layer with changing opacity, Drecki (2002) argued to integrate transparency in the data, meaning a higher uncertainty with increasing transparency.

In addition to the described abstract signifiers that only vary in a single visual variable, iconic symbols make use of metaphors, such as a clock for depicting time, and might be easier

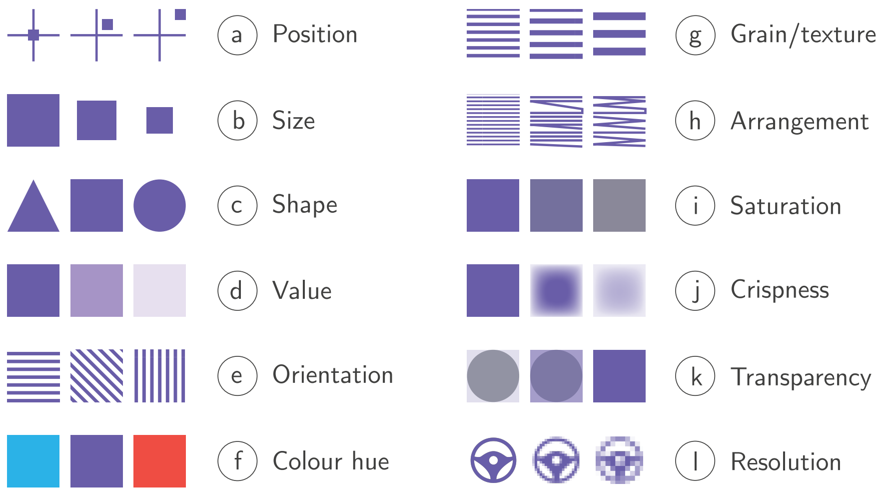


Figure 2.17: Extended set of visual variables, increasing uncertainty from left to right (adapted from MacEachren, Roth et al. 2012)

to match with different uncertainty components (MacEachren, Roth et al. 2012). However, abstract signifiers utilise preattentive visual processes and might therefore be quicker to identify than the more complex iconic sign vehicles (see Subsection 2.4.3) (MacEachren 2004; MacEachren, Roth et al. 2012). MacEachren, Roth et al. (2012) conducted two linked empirical studies to examine the intuitiveness of the presented abstract as well as additional iconic variables for communicating the outlined different uncertainty components in a geographic context (Thomson et al. 2005). The results indicate that particularly crispness and location can present uncertainty very intuitively, followed by arrangement, value, size, and transparency. In accordance with expectations, abstract vehicles led to quicker judgements, while iconic symbols were more accurately judged.

2.8.2.3 Anthropomorphic Uncertainty Communication

Anthropomorphic features are a suitable means for promoting trust in automation, rendering them a compelling uncertainty communication method (Visser, Krueger et al. 2012; Hoff and Bashir 2015; Pak et al. 2012; Waytz, Heafner and Epley 2014). Anthropomorphising interfaces implies the incorporation of human-like features, for instance human voices or facial expressions. The interaction of humans with automation interfaces that resemble basic human characteristics invoke similar trust-related behavioural outcomes as the interaction with other humans (Visser, Monfort et al. 2017). Further, observations of drivers' interaction with in-vehicle speech systems highlighted a behavioural tendency to treat somewhat human-like machines as if they were true humans. Likely triggered by the presence of a human voice, participants assigned further human-like characteristics such as gender or personality to the speech system and made it responsible for mistakes (Large and Burnett 2019). Anthropomorphic features were also shown to increase trust resilience, thus preventing disuse of automation. Following system failures, users interacting with anthropomorphic interfaces showed smaller trust declines than those using systems without human-like features, potentially the result of different expectations regarding the reliability of machines and humans (Visser, Krueger et al. 2012).

The use of anthropomorphic uncertainty interfaces therefore promises to dampen trust declines following system failures, contributing to trust repair (Visser, Pak and Shaw 2018) (see Subsection 2.7.1). However, a discrepancy between uncertainty communication must be noted. Whereas the paramount objective of uncertainty communication is to calibrate trust and thus to likely reduce trust in automation, anthropomorphism unilaterally increases trust, potentially leading to an imbalance which impedes accurate trust calibration. Suchlike ramifications must be considered when using anthropomorphic features to promote trust repair and trust in general.

2.9 Research Opportunities and Questions

This section consolidates the presented analysis of the literature and deduces specific research focus areas that address the formulated research questions (see Section 1.2). The transparency of automated systems was shown to affect the formation of dynamic trust and was identified as a major influencer of the trust calibration process (see Section 2.7). Numerous publications within the automotive as well as the military domain have indicated that the communication of uncertainty can positively influence the trust formation process and appropriate reliance on automation by increasing the transparency of the system. With inappropriate trust and insufficient transparency being two of three main contributors to the OOTL performance problem (see Subsection 2.6.2.2), the communication of uncertainties is expected to mitigate the safety concerns that are associated with automated driving systems (see Section 2.6). Addressing gaps in existing research (see Subsection 2.8.2.1), this project focuses on the investigation of system uncertainties in the context of automated driving. In the following, the identified research opportunities are further specified, resulting in overarching research questions (see Table 1.1 for an overview).

2.9.1 Implications of Dynamic Uncertainty Communication

Research relating to uncertainty communication in a driving context and other domains such as aviation has highlighted the benefits of uncertainty communication for trust calibration and task performance (see Subsection 2.8.2). Considering the various interrelations between trust calibration and other concepts such as workload or SA (see Figure 2.14), a more comprehensive investigation of the implications of uncertainty communication is necessary that takes all closely related factors, and not only trust and performance, into account.

The existing approaches to automotive uncertainty communication require drivers to regularly glance to the instrument cluster to perceive changes. This poses several significant problems. First, even if operators employ an optimal attention allocation strategy, changes may be missed (see Subsection 2.6.2.2). Second, the demand to regularly change focus between two spatially separated AOIs within the eye or head field (glance from centre console to instrument cluster or from road to instrument cluster) increases information access effort (see Figure 2.7) and thereby likely the perceived workload of users.

Besides the expected HF issues of using an uncertainty display located in the instrument cluster, the knowledge of the system's fallibility may lead to a feeling of uneasiness. Finally, the impact of the resolution of uncertainty information has yet to be explored. While Helldin et al. (2013) provided uncertainty information using seven levels, the behaviour associated with each level has not been explored.

In a first step, it will be investigated whether presenting uncertainty content in the instrument cluster has a negative impact on workload, gaze distribution, and measures of uneasiness. This also involves an analysis of user behaviour relative to different uncertainty levels. The following research question can be formulated:

RQ 1: What are the implications of dynamic uncertainty communication on key parameters of human-automation interaction?

Chapter 4 outlines the corresponding empirical study and discusses the impact of uncertainty communication on each dependent variable.

2.9.2 Unobtrusive Communication of System-Wide Uncertainty

As presented in Subsection 2.5.2, the practicability of NDRTs is a key factor for the success of vehicles equipped with automated driving systems. Approaches to uncertainty communication that impede the practicability of NDRTs should therefore be avoided. As previously noted, the proposed approaches by Beller, Heesen and Vollrath (2013) and Helldin et al. (2013) require users to regularly shift their attention to the instrument cluster. While this has already indicated benefits for the practicability of NDRTs, users' trust calibration and the manual driving performance following takeovers relative to systems without an uncertainty display, the designs can be improved to consider the usage context. Specifically, Subsection 2.5.2 has highlighted that driving cannot be considered the primary task with activated automation. Rather, NDRTs will assume this role. Thus, users can be expected to be fully engaged in tasks other than driving. Presenting crucial information about the system's state in the instrument cluster using a digital display impedes the perception of uncertainty information and thus reduces the benefits of uncertainty communication. Thus, an approach is required that allows users to perceive the uncertainty information while being engaged in NDRTs. Hence, a display that affords users to be *peripherally aware* of its content while the user is attending another AOI with focal attention is needed. Suchlike displays are hereafter referred to as *peripheral awareness displays* (see Section 5.1).

Chapters 5 to 7 explore and apply a development process for peripheral awareness displays and compare the developed display to a digital display located in the instrument cluster. The following broad research questions will be addressed:

RQ 2: How can the systematic evaluation of peripheral awareness displays be ensured?

RQ 3: How can uncertainty information be intuitively conveyed using peripheral awareness displays?

RQ 4: What are the implications of peripheral uncertainty communication on key parameters of human-automation interaction relative to the use of digital displays?

2.9.3 Communication of Function-Specific Uncertainties

Additionally, the analysis outlined in Subsection 2.8.2.1 revealed that the existing publications do not address the dimension of functional detail in the trust calibration process (see Section 2.7).

Both Beller, Heesen and Vollrath (2013) and Helldin et al. (2013) communicated system-wide uncertainties. The proposed interfaces are therefore characterised by a minimum functional specificity (see Section 2.2). It is hypothesised that the provision of function-specific uncertainties can help users to understand which part of the DDT is affected and may thus counteract the OOTL performance problem. While no related studies have been identified in a driving context, knowledge can be transferred from other domains. As outlined in Subsection 2.8.2.2, the integrated display of uncertainties with the affected data can potentially improve the interaction process by leading to a more appropriate reliance on automation (see Subsection 2.8.2.2) (Neyedli, Hollands and Jamieson 2011). Additionally, the use of graphics has been shown to be at least equally effective as numerical or verbal communication. Combined with the benefits outlined in Subsection 2.5.3.3, this favours caHUDs and HMDs for the communication of function-specific uncertainties. Since caHUDs do not require additional equipment provided by the user, they will be preferred over HMDs within this project. Visual variables originating from GIScience can be used as basic blocks for building these interfaces (see Subsection 2.8.2.2). Within this research focus area it must be identified how uncertainties can be communicated regarding the specific DDT functions, i. e. lateral vehicle control, longitudinal vehicle control, and OEDR (see Subsection 2.5.2).

As schematised in Figure 2.8, the function OEDR immediately affects the lateral as well as longitudinal motion control. Thus, uncertainties in the former inevitably affect the latter. To prevent uncertainties from being communicated twice and to avoid potential confusions, two new terms are introduced. This thesis expands the meaning of longitudinal and lateral vehicle control to include all aspects of the OEDR that influence them. To prevent ambiguities, the labels *lateral vehicle motion* and *longitudinal vehicle motion* are hence used:

- **lateral vehicle motion**: includes all processes that affect the motion of the vehicle in lateral direction; this encompasses, for example, the detection of lanes or the identification of traffic signs that indicate a necessary lane change
- **longitudinal vehicle motion**: includes all processes that affect the motion of the vehicle in longitudinal direction; this encompasses, for example, the detection of preceding vehicles and the maintenance of an appropriate gap or the identification of traffic signs that indicate speed limits

In a first step, the lack of research with respect to the impact of changes in the functional specificity of uncertainty information prompts exploratory research efforts using qualitative methods to supplement quantitative data (see Subsection 3.2.4). Further, the previously outlined visual variables can be investigated in terms of their suitability for conveying different uncertainty levels using caHUDs (or: AR HUDs). This leads to the following general research questions that are addressed in Chapters 8 and 9:

RQ 5: What are the qualitative and quantitative implications of increasing the functional detail of uncertainty information?

RQ 6: What visualisation layouts and variables are best suited for conveying uncertainties with a higher functional detail?

2.9.4 Proposed Interaction Concept

To combine the potential benefits of the unobtrusive and function-specific communication of uncertainties, a two-step process is proposed (see Figure 2.18). First, the attention of the user is shifted towards the FRD through an initially unobtrusive stimulus that becomes more salient with increasing uncertainty. In contrast to previous solutions, this allows the user to remain completely engaged in NDRTs as long as the system is fully capable of handling the current situation. Once users shift their attention towards the FRD as a result of an increased salience, function-specific uncertainties will be available within the FRD (step two). It is anticipated that this will allow users to quickly assess which aspect of the DDT is affected and will support them in localising the failure, thus counteracting the OOTL performance problem and supporting the build-up of an appropriate mental model.

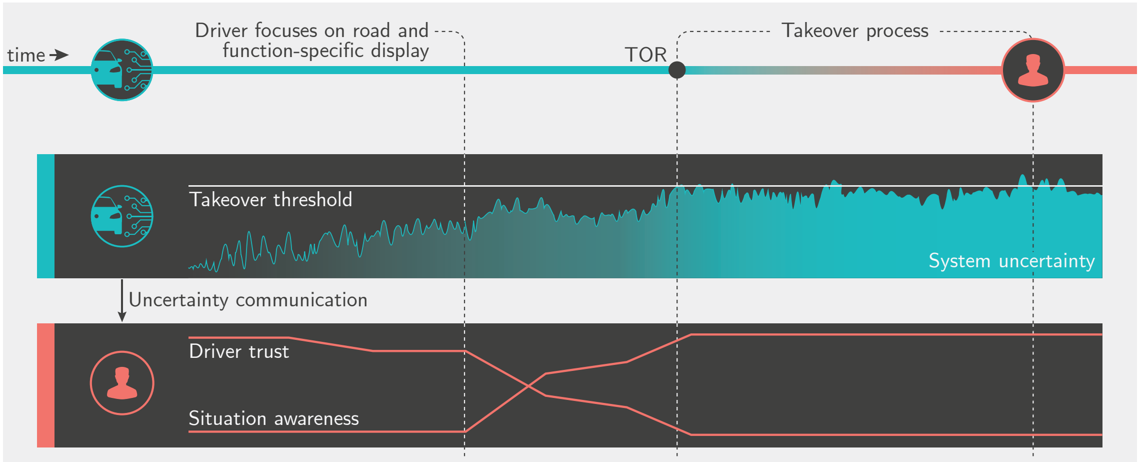


Figure 2.18: Schematic depiction of the proposed two-step interaction process

CHAPTER 3

Research Methodology

This thesis includes data collected from three driving simulator studies, three user-based experiments, and a heuristic evaluation. Although each study was conducted independently, the common methodological features will be discussed in this chapter, specifics of each experiment are detailed in the relevant subsequent chapters.

First, an overview of the chosen approach is given that outlines how each research question was addressed with distinct but interconnected research stages. Second, specific research tools and methods are discussed as means for fulfilling the thesis aims (see Section 1.2). Finally, experimental designs and statistical methods are reviewed in terms of their suitability for extracting information from each selected research method.

3.1 Research Approach

The research presented in this thesis has taken a controlled stepwise approach to investigating the formulated research questions. Insights obtained from initial research efforts informed later work, creating a chain of interconnected experiments that converged towards fulfilling the thesis objectives. Existing research, as discussed in Chapter 2 and consolidated in Figure 2.14, served as a starting point.

The applied research approach is summarised in Figure 3.1. The first stage was aimed at laying the foundation for all experiments that succeeded it and addressed the research question formulated in Subsection 2.9.1. Specifically, it examined the hypothesised shortcomings of current approaches to uncertainty communication and highlighted aspects that showed room for improvement (see Chapter 4). Following stage 1, the research was split into two tracks, a division that also manifests itself in the thesis structure (see Figure 1.1). Track A targeted the research questions described in Subsection 2.9.2. As such, it aimed to identify an uncertainty communication method that allowed users to perceive the conveyed information without the need for visual focus. This track was subdivided into three stages, whereby each phase built immediately on the other. Stage A2 was required to establish a framework and process that would serve as a scientifically grounded basis for the development of a peripheral awareness display for uncertainty communication (see Chapter 5). The subsequent stage, A3, applied the developed framework and process (see Chapter 6). In the final stage of track A, the developed display was evaluated against the instrument

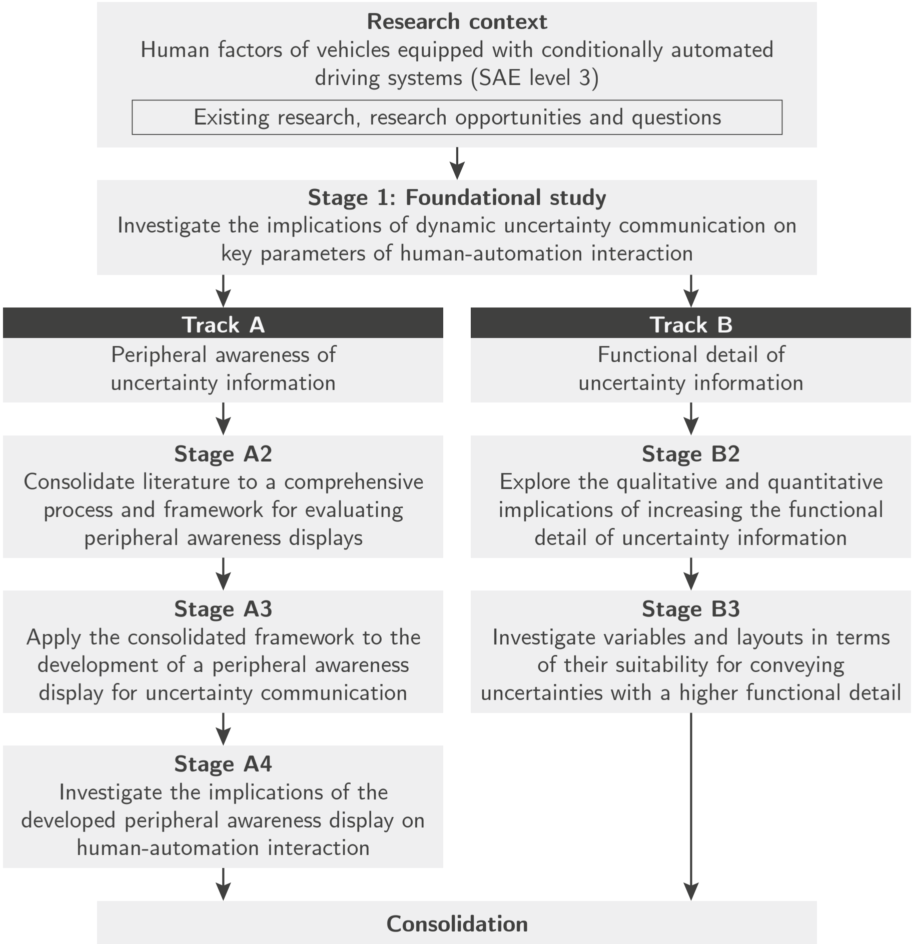


Figure 3.1: Overview of research approach

used in stage 1 to assess if the intended benefits for human-automation interaction were achieved (see Chapter 7).

Track B addressed the research questions formulated in Subsection 2.9.3. Consisting of two stages, the qualitative implications of increasing the functional detail of uncertainties were initially explored to gain general information about how and why this may or may not be beneficial for human-automation interaction (see Chapter 8). Continuing in this context, the second stage investigated how an increase in functional detail should be conveyed in order to be intuitively and effortlessly processed by users (see Chapter 9). The research approach concludes with a consolidation of the findings and an interpretation against the backdrop of the research context, i. e. the human factors of vehicles equipped with conditionally automated driving systems (SAE level 3, see Chapter 10). Throughout the duration of the project, the research efforts were accompanied and guided by literature surveys that ensured regular updates of related work.

3.2 Research Methods

The individual stages of the outlined research approach employed a range of different research methods from various disciplines. Besides methods originating in the automotive domain such as the use of driving simulators, the research methodology was derived from related disciplines like human-computer interaction. This section first outlines overarching principles regarding participant selection and pilot work that were applied in all experiments. Next, specific methods for data collection and analysis are discussed.

3.2.1 Sampling Frame and Design

The outlined research stages require experiments that involve human participants. For reasons of practical feasibility related to costs and time, a subset of individuals, i. e. a sample, was drawn from the overall target population. At the start of the research project, a sampling frame was defined as a representation for the population of interest (Shapiro 2008a). As such, all participants were required to hold a driving licence that legally allowed them to drive on public UK roads. To be able to perceive all conveyed information, participants were additionally required to have normal or corrected-to-normal acuity of vision. To reduce the impact of individual differences in psychomotor abilities, elderly populations (60+ years) were excluded (Vrtunski, Patterson and Hill 1984). Further, researchers involved in this project were also barred from all experiments.

For reasons of accessibility and efficiency, the population was limited to students and staff from Loughborough University and the sample design employed convenience sampling. While this is a commonly taken approach in psychological research (Battaglia 2008), the sampling method entails the problem of not having a known non-zero probability of selecting each element in the population. For instance, participants with certain attributes could unconsciously be selected while others are excluded (Shapiro 2008b). To counteract this issue, calls for participants were spread across all university departments to achieve a more even probability of selecting each element of the population. Further emphasis was placed on the recruitment of participants differing in age and profession to prevent the sole use of university students with high levels of digital proficiency. To avoid a skewed sample, participants were not financially reimbursed as this would disproportionately attract lower income groups such as students. Moreover, financial reimbursements were abstained from for ethical reasons. Specifically, the monetisation of participation would 'degrade the idea of a common good that research contributes to, and instead transform it into another marketised exchange (Head 2009, p. 343)'. Additionally, reuse of participants was avoided. Demographic data of the participants selected for each experiment are described in the respective sections of this thesis.

3.2.2 Pilot Work

The data collection for all experiments outlined within this thesis was preceded by pilot tests to identify any problems with the experimental design and setup. Thereby, the pilot test procedures mirrored the procedures employed in the respective experiments and included researchers experienced in experimental design as pilot participants (Rothgeb 2008). For each experiment, at least six participants completed the pilot tests.

Several optimisations were implemented as a consequence of pilot testing. For instance, pilot testing revealed that the light intensity in the driving simulator laboratory was not sufficient for eye tracking when the ceiling lights were turned off. Further, the randomisation of vehicle models in the simulation scenario led to the inclusion of salient models such as a DMC DeLorean, which distracted participants. Consequently, the ceiling lights remained activated and the simulated vehicle types were limited to generic shapes, such as sedans or SUVs, that mainly varied in colour. Further consequences of pilot testing are detailed in the corresponding sections of this thesis.

3.2.3 Quantitative Data Collection and Analysis

Various quantitative research tools and data collection methods were employed within this research project. Hereafter, a selection of key methods is discussed in conjunction with their corresponding metrics.

3.2.3.1 Driving Simulator Data

Experiments that required a dynamic driving environment were conducted in a fixed-based driving simulator running STISIM Drive 3 (Systems Technology 2017) that consisted of a mock-up Land Rover cockpit and three projectors creating a combined 135° field of view (see Figures 3.2 and 6.6).

The use of driving simulators has several advantages over on-road driving conditions. As such, it enables the simulation of technologies that are not yet available on public roads, for instance driving automation of SAE level 3. Further, hazardous situations such as emergency takeovers can be investigated without exposing participants and other road users to physical risk (Kaptein, Theeuwes and Van Der Horst 2007). Moreover, simulations allow full experimental control, thus ensuring equal conditions for all participants and avoiding unintended influences of other factors such as weather or traffic density. Additional benefits include a reduced ecological footprint and lower costs (Kaptein, Theeuwes and Van Der Horst 2007).

However, questions of validity arise when using simulated instead of real-world environments. In general, the validity relating to a specific research question ‘refers to the degree

Figure 3.2: Driving simulator setup

that the simulator evokes the same behaviour as would be shown in reality under similar circumstances (Kaptein, Theeuwes and Van Der Horst 2007, p. 31)'. In this context, internal validity refers to the extent of causality between the change of an independent variable and the measured effect on a dependent variable. External validity refers to the degree to which the obtained results can be generalised for wider populations. Relative validity indicates to what degree the relative effect size and direction correlate with real situations, whereas absolute validity requires the absolute effect sizes to coincide. Prior work has investigated and confirmed the relative validity of driving simulators as research tools. Eriksson, Banks and Stanton (2017), for instance, recorded a statistically significant correlation between the control transition time in a driving simulator and an on-road driving condition with an SAE level 2 automated driving system (Kaptein, Theeuwes and Van Der Horst 2007).

Several measures were taken to ensure high levels of internal and relative validity in the experiments outlined in this thesis. As such, high fidelity input and output devices comparable with those installed in commercial vehicles were used and the simulation image was presented using high resolution projectors that generated a large field of view. Further, only the independent variables stated in the experiment descriptions were varied, thus minimising the impact of other influences such as differences in traffic density.

The experiments are, however, limited in terms of their absolute and external validity. To the knowledge of the author, there is currently no real-world data concerning higher levels of automation (SAE level 3+) available with which the driving simulator data can be compared. Consequently, general questions of validity concerning the use of driving simulators remain. Due to a lack of real risk to the well-being of participants, reaction times are expected to decrease in on-road situations (Eriksson, Banks and Stanton 2017). As a consequence of using a fixed-base simulator, vehicle dynamics were not conveyed,

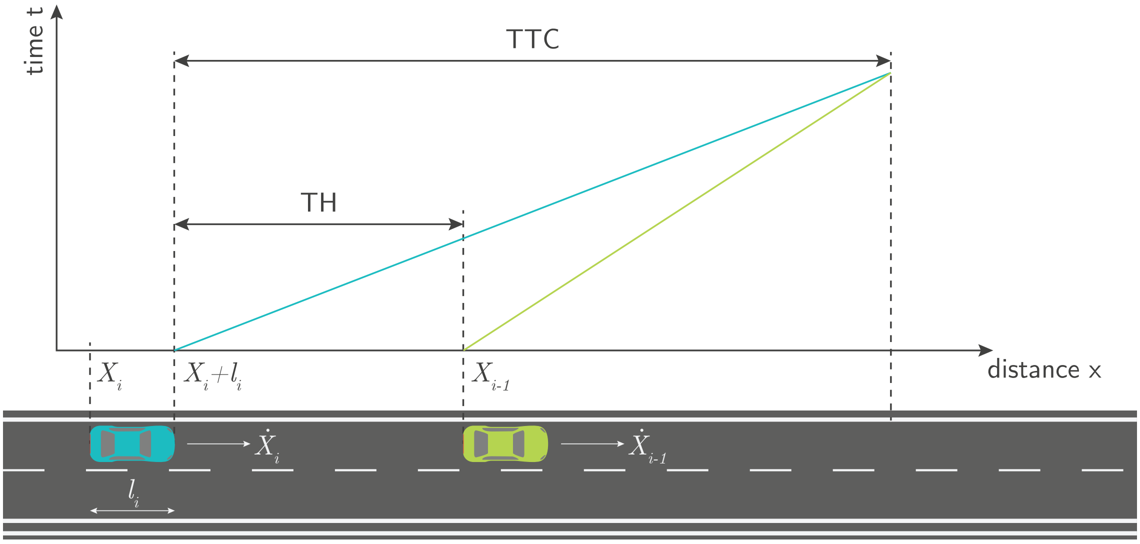


Figure 3.3: Schematic depiction of time headway and time to collision (adapted from Bubb et al. 2015; Vogel 2003; Minderhoud and Bovy 2001)

thus impeding participants' sense of acceleration (Bellem et al. 2018). Further limitations include the focus on a particular road type, i. e. motorways, and the exclusion of elderly drivers, thus limiting external validity.

Besides replicating the intended usage scenario, the driving simulator served as a means for data collection. Specifically, metrics relating to takeover performance and quality were extracted. The indicators of takeover quality can be derived from metrics used to assess manual driving quality, particularly lateral acceleration (Bubb et al. 2015). To be considered comfortable, lateral acceleration values should not exceed $0.3\,\mathrm{m\,s^{-2}}$ (Bubb et al. 2015, p. 32). Time headway (TH) and TTC can be regarded as standard measures for driving safety (Vogel 2003; Green 2013; SAE International 2015). TH is defined as the time that passes between ego and lead vehicle reaching the same location (see Figure 3.3) (Vogel 2003):

$$TH = t_i - t_{i-1} \tag{3.1}$$

with TH = time headway
t_i = time at which ego vehicle i passes a certain location
t_{i-1} = time at which the leading vehicle passes the same location

On the basis of real-world traffic flow measures, Vogel (2003) recommends the use of TH for enforcement purposes, whereas TTC should be used to evaluate the safety of specific traffic situations. As the evaluation of (takeover) safety is a major parameter of interest, TTC was used instead of TH. TTC is defined as 'the time span left before two vehicles collide if nobody takes evasive action (Vogel 2003)', see Equation 3.2 (Minderhoud and Bovy 2001).

$$TTC_i = \frac{X_{i-1}(t) - X_i(t) - l_i}{\dot{X}_i(t) - \dot{X}_{i-1}(t)} \quad \forall \dot{X}_i(t) > \dot{X}_{i-1}(t) \tag{3.2}$$

with $\dot{X}_i$ = speed of ego vehicle i $\quad$ $\dot{X}_{i-1}$ = speed of lead vehicle $i-1$
X_i = position of vehicle i $\quad$ X_{i-1} = position of lead vehicle $i-1$
l_i = length of vehicle i

Thereby, particularly minimum time to collision (MTTC) values are relevant when evaluating takeover performance (Beller, Heesen and Vollrath 2013). MTTC can be defined as the minimum TTC for which a continuation with the current speed and heading direction would result in a collision. STISIM 3 provides MTTC values relative to each nearby vehicle as a predefined parameter. However, the provided values were not valid when a standing vehicle was passed by the ego vehicle on another lane. Specifically, an MTTC of 0 s was recorded for these situations, indicative of a crash. To counteract this, the recorded driver input and/or vehicle position parameters were used in conjunction with recorded TTC values to approximate MTTC in such situations. For manoeuvres in which a collision with a lead vehicle was avoided by changing lanes, the recorded TTC at the point of lane change initiation was used as MTTC.

Besides TTC, time to takeover (TTT) can be regarded as a further safety indicator for takeovers (Zeeb, Buchner and Schrauf 2015). TTT was measured as the time span between a TOR and the first driver input to either steering wheel ($|\varphi| \geq 3°$), brake, or acceleration pedal (Helldin et al. 2013). Thus, TTT is synonymous with reaction time of drivers following TORs. In preparation for the takeover reaction, drivers typically gaze towards the FRD and place a hand on the steering wheel, referred to as *readiness to drive* (Zeeb, Buchner and Schrauf 2015; Large, Burnett, Salanitri et al. 2019). The time to driver readiness was not reported in experiments as it is mostly reflexive and, consequently, not influenced by aspects such as visual distraction (Zeeb, Buchner and Schrauf 2015). In contrast, the measure TTT includes the time taken to cognitively process information and is therefore expected to be more significantly affected by uncertainty communication.

3.2.3.2 Eye Tracking Data

Eye tracking analysis relies on the premise that visual fixations (see Subsection 2.4.3.2) predict attention allocation (Tobii AB 2019a). Phenomena such as change blindness and concepts like covert attention highlight that this assumption is not always fulfilled (see Subsection 2.4.3.1). Despite its limitations, eye tracking analysis provides unbiased, real-time quantitative insights about operator behaviour without being obtrusive (Tobii AB 2019b).

Figure 3.4: Definition of areas of interest

Prior research has indicated an inverse relation between the monitoring activity, i. e. fixations, towards the FRD and trust of operators in automation (Muir and Moray 1996; Hergeth et al. 2016) (see Subsection 2.7.1). Further, measures such as the off-road glance time serve as indicators for manual driving safety (Seppelt, Seaman et al. 2017). In reference to related work (Hergeth et al. 2016), eye tracking data were analysed using monitoring frequency and ratio. Monitoring frequency describes how often participants checked the behaviour of the automated driving system and the driving environment, see Equation 3.3.

$$\text{Monitoring frequency} = \frac{n_{\text{monitoring glances}}}{t_{\text{section}}} \tag{3.3}$$

Figure 3.4 illustrates the AOIs defined for the experiments within this research project, whereby the AOI *uncertainty display* depended on the display type that is being used. The variable $t_{section}$ refers to the duration of the segment for which the monitoring data is analysed. Monitoring glances were thereby defined as eye movements from non-driving-related to driving-related AOIs, namely from the NDRT to the FRD, instrument cluster, or the uncertainty display. Consecutive fixations on the same AOI separated by blinks of less than 120 ms were combined and fixations not exceeding 120 ms were omitted from evaluation (Hergeth et al. 2016; Jacob and Karn 2003). The relative accumulated duration of fixations regarding a particular AOI was assessed using the monitoring ratio, see Equation 3.4.

$$\text{Monitoring ratio} = \frac{\Sigma_n^{i=1} t_{\text{fixation } i}}{t_{\text{section}}} \tag{3.4}$$

Other types of eye movement such as saccades (see Subsection 2.4.3.2) are not included in the monitoring ratio. Hence, the accumulated monitoring ratios of all recorded AOIs do not add up to 100 %. As the durations of fixations and saccades depend on the task at hand (Rayner 2009), absolute monitoring ratios are not comparable across different contexts.

The wearable eye tracker Tobii Glasses 2 was used to capture the viewing behaviour of participants. In accordance with the corresponding manual (Tobii AB 2017), each recording was preceded by a calibration process.

3.2.3.3 Situation Awareness

Insufficient levels of SA are a major contributing factor to the OOTL performance problem and have been shown to impede safe takeovers (see Subsection 2.6.2). Consequently, enhancing operator SA is a primary objective for any HMI design effort aiming to improve human-automation interaction in the context of automated driving. Therefore, SA is a relevant metric for assessing system designs that requires reliable and valid measurement techniques.

The most frequently used and validated tool for measuring SA as defined within this research project (see Subsection 2.6.2.2) is the Situation Awareness Global Assessment Technique (SAGAT) proposed by Endsley (1995a) (Salmon, Stanton, Walker and Green 2006; Endsley and Garland 2000). Employing SAGAT, the simulation is temporarily interrupted and all displays blanked without prior notice while participants are prompted to quickly communicate their perception of the situation. Participant responses are then evaluated against a ground truth.

This contrasts alternative methods such as the Situation Awareness Rating Technique (SART) (Taylor 1990) which infer SA from subjective participant responses to questionnaire items following the experimental scenario. While this does not require an interruption of the experimental scenario, Salmon, Stanton, Walker, Jenkins et al. (2009) highlighted that the validity of SART for measuring participant SA is questionable as it was shown to be more closely linked to operator workload than SA-related measures.

A number of additional tools for measuring SA exist, for instance the Systematic Human Error Reduction and Prediction Approach (SHERPA) (Hughes et al. 2015). The experiments within this research project apply SAGAT as its validity is supported by extensive evidence and it is widely accepted within the research community (Salmon, Stanton, Walker and Green 2006).

Besides measures that rely on explicit participant responses, more implicit indicators for SA exist. As such, the mirror-checking behaviour of drivers can be assessed (Large, Burnett, Salanitri et al. 2019). The driving simulator implementation employed within this research project, however, impeded the use of this measure. Specifically, the mirrors were all rendered as part of the virtual driving scene, thus removing the spatial separation between AOIs found in real vehicle cockpits (see Figure 3.2). Particularly the right side mirror, which is arguably most significant for the implemented takeover scenarios, was thereby positioned within the immediate FRD. The accuracy of discrimination between glances towards the road and right side mirror checks was found to be insufficient as a reliable research measure. Consequently, glances to the FRD in general (including mirrors) were used as an implicit measure of SA.

3.2.3.4 Trust Measures

As illustrated in Figure 2.14, the users' trust in automation indirectly affects the acquisition and maintenance of SA. Several studies point towards using the user's gaze behaviour as an indication for trust (see Section 2.7) (Hergeth et al. 2016). Nonetheless, eye tracking poses the problem of looking but not seeing, meaning that the gaze of participants is pointed in a particular direction but they are not cognitively focusing (Parasuraman and Manzey 2010; Healey and Enns 2012). Consequently, a number of complementary methods should be used to measure trust. The majority of publications measuring trust use questionnaires to assess the belief of participants in the trustworthiness of an automated system. Jian, Bisantz and Drury (2000) propose a scale of trust consisting of twelve individual questionnaire items. While this allows for the consideration of different trust facets, it impedes the repeated measurement in dynamic driving simulator studies due to its length. The questionnaire is therefore not suitable to measure dynamic changes in trust. Alternatively, single-item trust scales have been proposed that prompt participants to indicate their current trust in automation on a percentage scale (Hergeth et al. 2016; Beller, Heesen and Vollrath 2013; Ginwalla et al. 2012). This method allows the repeated measurement of overall subjective trust without the interruption associated with elaborate questionnaires. It is debatable, however, if expressed beliefs translate to real-world behaviours. Miller, Johns et al. (2016) proposed the *trust fall* to investigate the actual behaviour of humans in safety-critical situations. The trust fall refers to a team-building exercise in which a participant expresses his or her trust in another participant by falling backwards into the arms of the latter. Transferred to driving, they presented participants with a short gap lane incursion while the automated system was engaged. Intervening would display a lack of trust in the capabilities of the system while no intervention would indicate a high trust in the system. The delimitations of the simulation platform used in the experiments outlined in this thesis (see Subsection 3.2.3.1) restricted driver-initiated transitions of control and therefore impeded the use of the trust fall as a method for measuring operator trust in automation.

Alternatively, the solving rate of NDRTs can be used as an objective indication for trust. With the monitoring frequency towards driving-related AOIs and trust being inversely

(a) Target present **(b)** Target present **(c)** No target present

Figure 3.5: Visual search task employed in the experiments of Chapters 4 and 7. For (a) and (b), participants were required to tap on the left side of the touchscreen. For (c), a tap to the right of the screen's vertical centre line was needed.

related (Muir and Moray 1996; Hergeth et al. 2016), participants can be expected to increasingly engage in NDRTs with higher trust. Prior research has used visual search tasks for this application (Beller, Heesen and Vollrath 2013) as it is one of the most significant cognitive subtasks in the context of driving (Shinohara et al. 2010). The visual search task prompts participants to identify if a predefined target is present among a collection of distractors (Treisman and Gelade 1980). Within the experiments outlined in this thesis, participants needed to identify whether a blue letter or an S is present and indicate their answer with a tap to either the left side (target present) or right side (no target present) of the touchscreen (see Figure 3.5). In addition, physiological measures such as heart rate may indicate differences in trust (Waytz, Heafner and Epley 2014), the measurement of which is discussed in Subsection 3.2.3.2.

3.2.3.5 Workload

Being affected by both automation characteristics and the presence of NDRTs, operator workload itself has an impact on top-down attention allocation and thereby influences human-automation interaction (see Figure 2.14), thus prompting its inclusion as a dependent variable. Moreover, the workload associated with a particular interaction affects usability (Bevan and Macleod 1994). With more than 550 studies that applied or reviewed the tool, the NASA Task Load Index (NASA-TLX) is the most frequently used scale for estimating the workload of operators during or immediately after task performance (Hart 2006; Salmon, Stanton, Walker and Green 2006). The NASA-TLX was developed by Hart and Staveland (1988) and consists of six dimensions (see Table 3.1): mental demand, physical demand, temporal demand, effort, performance, and frustration level. Participants are asked to rank each dimension within given endpoints, with the combination of rankings leading to an overall workload metric. In reference to the Yerkes-Dodson law (Yerkes and Dodson 1908),

Table 3.1: NASA-TLX rating scale and definition (Hart 2006)

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

medium levels of arousal lead to best task performance. Similarly, scores around the medium range of the workload scale are considered optimal (Stanton, Dunoyer and Leatherland 2011). In spite of its popularity, the application of the NASA-TLX is characterised by a number of limitations. As participants fill in the questionnaire following the experimental task, the validity of the tool relies on the ability of participants to accurately recall their mental and physical states during the experiment. Moreover, recent studies questioned the construct validity of the NASA-TLX questionnaire (McKendrick and Cherry 2018). To counteract the former limitation, the NASA-TLX questionnaires were handed to participants immediately following the completion of experimental tasks.

3.2.3.6 User Experience

The previously outlined metrics mainly serve as safety indicators. However, aspects such as comfort or aesthetics must not be neglected when designing interfaces for commercial vehicles as they constitute relevant determinants for purchase decisions (see Figure A.5).

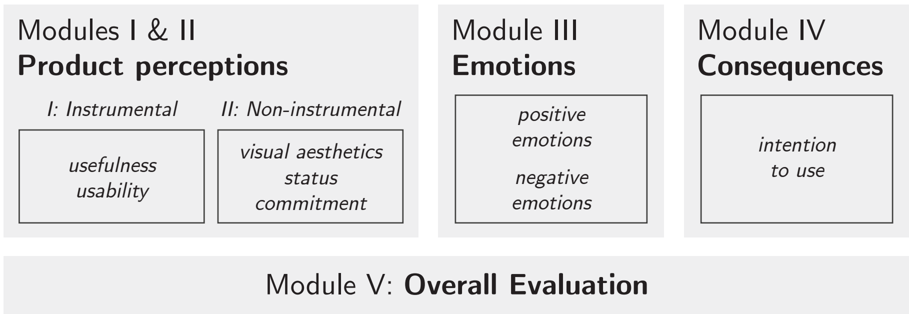


Figure 3.6: Structure of the meCUE questionnaire (Minge and Thuering 2016; Minge 2018a,b)

User experience can be regarded as an indicator for both pragmatic and hedonic system qualities (see Section 2.3). Several methods for measuring user experience (UX) exist, most of which utilise questionnaires (Pettersson et al. 2018). The most frequently applied questionnaires for measuring UX are AttrakDiff (Hassenzahl, Burmester and Koller 2003) and NASA-TLX (Pettersson et al. 2018). However, alternative questionnaire designs allow a more comprehensive approach despite fewer applications. The meCUE questionnaire proposed by Minge, Thüring and Wagner (2016) and Minge and Thuering (2016) can be regarded as an advancement of previous methods, such as AttrakDiff, as it employs a more comprehensive approach based on the CUE model presented in Section 2.3. Besides considering pragmatic and hedonic qualities, it also includes an assessment of emotional reactions. The standardised meCUE questionnaire consists of 34 items that can be split into five different modules (see Figure 3.6). The underlying CUE model also accounts for the bi-directional relationship between emotions and instrumental as well as non-instrumental qualities (Aranyi and Schaik 2016). As such, the qualities of a system do not only affect the emotions in a uni-directional way (see Figure 2.4), but also the other way around. UX measurements taken within the outlined experiments were based on the meCUE questionnaire (see Chapter 5).

As highlighted by Nass and Moon (2000), humans tend to avoid giving unpleasant answers to both humans an inanimate agents if the critique is addressed at them. For instance, a computer that prompts users to assess its own performance will receive more favourable responses compared to the assessment on another computer. Consequently, all evaluations that required participants to assess system characteristics, including UX assessments, used external devices for data collection to prevent politeness effects.

3.2.3.7 Physiological Measures

It might appear that increasing levels of automation relieve drivers of the stress associated with driving. As discussed in Subsection 2.6.2.2, however, automation does not replace

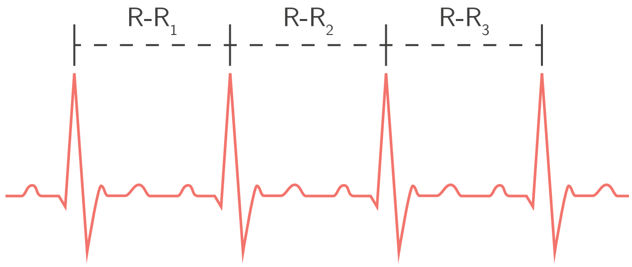


Figure 3.7: R-R interval

human performance, but rather changes it. Instead of actively controlling vehicle movement, users are required to remain vigilant in order to compensate for potential system failures. This sustained monitoring of automated driving systems and the unease associated with yielding control to an inanimate agent was shown to increase mental strain (Greenlee, DeLucia and Newton 2018; Morris, Erno and Pilcher 2017). Thus, increasing driving automation does not in itself lead to a reduction in stress. Knowledge of the inherent system uncertainties may further increase feelings of uneasiness, particularly when the uncertainty level implies an imminent system failure. Operator stress can therefore be regarded as a further indicator for the implications of a particular system design.

The experience of stressful situations stimulates a series of involuntary physiological responses that can serve as objective indicators for operator stress. The hypothalamus, an area of the brain, prompts the adrenal glands to release the hormone epinephrine into the bloodstream, leading to an observable increase in heart rate, respiratory rate, blood pressure, and blood sugar (glucose). As a second response, cortisol levels are increased to sustain the state of high alert (Havard Health Publishing 2011).

Each of the physiological responses can be measured as an indicator for stress. Non-invasive data collection methods are thereby preferred to reduce the health risks for participants. Being measurable with optical sensors, the heart rate of participants can serve as an initial indicator for stress (Havard Health Publishing 2011). The experiment outlined in Chapter 4 analyses the impact of uncertainty communication and takeovers on both absolute and normalised operator heart rate. For normalisation, feature scaling was applied to rescale the data (see Equation 3.5).

$$\text{bpm}_{normalised} = \frac{\text{bpm} - \text{bpm}_{min}}{\text{bpm}_{max} - \text{bpm}_{min}} \tag{3.5}$$

Besides heart rate, the variation in temporal intervals between heartbeats, referred to as heart rate variability (HRV), can serve as a further indicator for stress (Shaffer and Ginsberg 2017). For a healthy heart, the interval between adjacent heartbeats is not steady but shows complex, non-linear oscillations. Too little variation is an indicator for stress

(McCraty and Shaffer 2015). HRV is commonly calculated as the root mean square of successive differences (RMSSD) between R-R intervals measured with an electrocardiogram (ECG) (see Figure 3.7 and Equation 3.5).

$$\mathrm{RMSSD} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(\mathrm{RR}_i - \mathrm{RR}_{i+1})^2} \tag{3.6}$$

O'Neal et al. (2016) specified reference values for short measurements (10 s), $M \pm SD = 27.3\,\mathrm{ms} \pm 22.2\,\mathrm{ms}$. HRV is employed as a non-invasive indicator for stress in the experiment described in Chapter 7. The heart and respiratory rate of participants were measured as additional indicators for stress in the same experiment.

3.2.4 Qualitative Data Collection and Analysis

Qualitative research and analysis methods were used to supplement quantitative investigations and to collect exploratory data.

3.2.4.1 Qualitative Research Interview

The qualitative research interview is a conversational practice through which insights are gained by verbally interacting with a participant, i. e. the interviewee, or a group of participants (see, for instance, Chapter 8). In contrast to conversations in everyday life, qualitative research interviews are carried out to advance the researcher's knowledge of a specific topic and are often implemented as a *one-way dialogue* whereby the interviewee speaks most (Brinkmann 2008). Interview guides outline the content that is to be covered as part of an interview. The degree of specification detail lies on a continuum ranging from ethnographic interviews (least structured) to survey questionnaires (highly structured) (Morgan and Guevara 2008). Ethnographic interviews often start with *grand tour* questions that aim at developing a basic understanding of the interviewee's associations with the research topic, for instance 'Tell me about your typical approach to [...]'. These broad questions are followed up by *mini tour* questions that focus on receiving more detailed information about something the interviewee mentioned. Interview guides for survey questionnaire specify both the content of each question and the possible responses. In-between the outlined extremes in terms of prespecified structure lies the semi-structured interview. Thereby, the interview guide contains a framework of research questions, but the researcher is free to reorder questions and follow-up on answers that are of particular interest. This allows the researcher more control over the discussed topics as compared with ethnographic interviews and does not restrict the responses and information detail as with survey questionnaires (Ayres 2008).

Table 3.2: Key terms of a content analysis (Erlingsson and Brysiewicz 2017)

Term	Definition	Example
Theme	Expression of underlying meaning, latent content	*Not a person, just a body in the hectic emergency centre*
Category	Grouping of codes	*Staff actions and non-actions*
Code	Label, one or two words describing condensed form	*Left alone*
Condensation	Shortening of text while preserving core meaning	*Pushed to the middle of the room, walked away, left me*
Meaning unit	Text excerpt	*"They pushed me into the middle of the room and then walked away.. they just left me"*

Increasing abstraction

3.2.4.2 Content Analysis

Content analysis is a data analysis technique for the exploration and examination of recurring patterns in communication artefacts, e. g. interview transcripts (Bryman and Bell 2015). Its aim is 'to systematically transform a large amount of text into a highly organised and concise summary of key results (Erlingsson and Brysiewicz 2017)'. The systematic establishment of patterns allows the count of frequencies, thereby enabling the quantitative analysis of formerly qualitative data (Morgan 1993). Patterns in the data are discovered using *codes* – labels or keywords delineating each pattern. These codes can either be inferred from the data itself through careful reading (qualitative approach for small datasets) or are automatically applied by a computer (quantitative approach for large data sets) (Morgan 1993). Hsieh and Shannon (2005) present three approaches to content analysis and the coding of data. In *conventional content analysis*, coding categories are extracted directly from the data during data analysis. A *directed approach* derives the coding categories from a research hypothesis or prior findings before and during data analysis. The *summative approach* starts with a count of individual keywords and proceeds with a further analysis, for example regarding the context in which each word was used. Irrespective of the chosen approach, a similar analytical process containing seven stages is required for a qualitative content analysis (Kaid 1989): (1) formulation of research questions; (2) selection of a sample; (3) definition of categories; (4) outline of the coding process; (5) implementation of the coding process; (6) determination of the trustworthiness; (7) analysis of the results. Chapter 8 outlines how a content analysis was applied to interview responses.

3.2.5 Experimental Design and Statistical Methods

There are two primary options for the construction of experimental designs. Using a *between-subjects* design, each participant is exposed to one experimental treatment. In contrast, *within-subjects* designs expose participants to multiple treatments. Experiments that employ a combination of both designs utilise a *mixed* design. Each approach to experimental design has distinct advantages and disadvantages (Charness, Gneezy and Kuhn 2012). For within-subjects designs, internal validity does not depend on random participant allocation to rule out effects of individual differences in, for instance, task performance. Within-subjects designs generally require less participants than between-subjects designs and thus raise statistical power for similar sample sizes. However, within-subjects designs may prompt participants to act in a way that complies with experimenter's intentions, referred to as the *demand effect* (Rosenthal 1976; Charness, Gneezy and Kuhn 2012). Further, order effects must be avoided when employing within-subjects designs by balancing the number of participants exposed to each possible combination of treatments.

The experiments outlined in this thesis employed within-subjects designs when decisions between several equal alternatives were required. For instance, when different interfaces for the same application were compared (see Chapters 6 to 8). On the other hand, between-subjects designs were preferred when the intentions of the experimenter would be highly obvious to participants. For instance, when an uncertainty display was either present or absent. Consequently, the experiment outlined in Chapter 4 used a between-subjects design for the variable *uncertainty display*. Finally, the study presented in Chapter 9 employed a mixed design to allow the comparison of a large number of variables (within) while preventing that participants simply transfer the scores from one visualisation type to the other (between).

Statistical analyses followed the procedures described by Howell (2013). Generally, metrics that were measured once per participant and experimental condition were analysed using analysis of variance (ANOVA), t tests, or their nonparametric counterparts for non-normal distributions. The statistics were accompanied by measures for effect size (Cohen's d, η^2) where applicable. The specific methods used for each experiment are detailed in each corresponding section of this thesis.

Various metrics were measured multiple times for each participant and condition to indicate changes due to different treatments within each scenario. For instance, subjective trust levels were recorded several times for each participant and scenario (see Chapter 4). The repeated measurement results in non-independent data points that impede the use of prevalent statistical methods such as ANOVA. Specifically, the variance regarding each metric differs between participants, meaning that individual data points cannot be considered independent from each other. This can be bypassed by aggregating the data. However, the aggregation of non-independent data points would not take advantage of all collected data but rather use averages, thus limiting statistical power. Instead, linear mixed models (LMMs) (Snijders and Bosker 2012) were used to statistically analyse the data. For this, R (R Core Team 2019) and the package lme4 (Bates et al. 2018) were applied. In addition to linear regressions that take fixed effects and a general error term (ϵ), LMMs allow the specification of random effects, hence the term *mixed* model. As a consequence, LMMs computationally reduce the influence of individual participants or other random parameters on the results. However, valid effect sizes such as Cohen's d or η^2 are not readily available, particularly for multiple fixed effects (Brysbaert and Stevens 2018). The inclusion of all data and consideration of random effects was considered a greater benefit, effect sizes are only included in basic comparative tests, e. g. t tests.

As a prerequisite for the use of LMMs, homoscedasticity must be assumed. The assumption of homoscedasticity requires that the residuals for all predicted scores of dependent variables are approximately equal. To validate this assumption, residual plots must have a similar width for all predicted values (Tabachnick and Fidell 2013). The assumption of homoscedasticity was inspected for each analysis. The specific implementation of LMMs is detailed in each corresponding section.

CHAPTER 4

Implications of Dynamic Uncertainty Communication

Abstract This chapter presents the empirical evaluation conducted to investigate the implications of dynamic uncertainty communication on key parameters pertaining to human-automation interaction in the context of SAE level 3 (SAE International 2018). This level was selected due to the anticipated HF challenges it poses to human operators (see Subsection 2.5.6). Addressing RQ 1 (see Subsection 2.9.1 and Table 1.1), the effects of dynamic uncertainty communication on trust, takeover performance, workload, gaze distribution, and physiological measures were analysed under consideration of NDRTs. Founded in the integral model presented in Figure 2.14, the impact of dynamically conveying uncertainties is examined more comprehensively than in related studies (Beller, Heesen and Vollrath 2013; Helldin et al. 2013). To facilitate the comparability with previous work on uncertainty communication in the context of automated driving (Beller, Heesen and Vollrath 2013; Helldin et al. 2013), a visual display located in the instrument cluster was used to convey system uncertainties. The results of a driving simulator study demonstrate the chances and risks of communicating uncertainties in a driving context and thereby inform the development of a user-centred uncertainty display (see Chapters 5 to 7). While the results confirm the general benefits of uncertainty communication for trust calibration and SA, the use of a visual display required users to regularly shift their attention, resulting in increased workload. Aspects of this chapter, in part verbatim, are published in Kunze, Summerskill et al. (2019a).

Background	Objectives	Method	Results	Implications
Calibrate trust and attention allocation through uncertainty communication to improve SA and takeover performance. Existing research lacks holistic investigation.	Holistic evaluation of the implications of dynamic uncertainty communication on trust, monitoring behaviour, SA, workload, and takeover performance.	Driving simulator experiment using an anthropomorphic uncertainty display located in the instrument cluster. Participants engaged in visual search task.	More appropriate trust, gaze behaviour, and improved takeover performance when uncertainty display was available, but higher workload due to monitoring.	Dynamic uncertainty communication improves human-automation interaction. Uncertainty display in the instrument cluster leads to workload increments.

Figure 4.1: Overview of Chapter 4

4.1 Method

A driving simulator study was conducted to investigate the communication of uncertainties in a dynamic context that mirrors real-world scenarios. Based on the broad research opportunity specified in Subsection 2.9.1, the following experimental research questions were addressed (see Table 1.1):

1. How does dynamically communicating automation uncertainties affect operator trust, SA, and monitoring behaviour?
2. How does operator gaze behaviour relate to trust?
3. Does a higher resolution of uncertainty information affect operator trust and monitoring behaviour?
4. How does dynamically communicating automation uncertainties affect takeover performance and quality?
5. How does dynamically communicating automation uncertainties affect operator workload?
6. How does dynamically communicating automation uncertainties affect physiological measures?

4.1.1 Participants

A total of 34 participants (14 female) with an average age of 30.824 years (SD = 8.712 years, range = 19-52 years) volunteered to take part in the experiment. Participants held a driving licence for a mean time of 11.176 years (SD = 8.802 years, range: 1-36 years) and reported to drive an average of 6248.529 miles per year (SD = 5122.152 miles per year). Participation was exclusively self-motivated. Participants were not financially reimbursed for completing the study and there were no external incentives for better performance (see Subsection 3.2.1). Ethical approval was granted by the Ethics Approvals Sub-Committee of Loughborough University.

4.1.2 Design

The research questions were investigated using a 2 (uncertainty display, between-subjects) $\times$ 4 (visibility range, within-subjects) mixed experimental design. The factor uncertainty display was implemented using two levels (1: uncertainty display unavailable; 2: uncertainty display available). Results of t tests indicated non-significant differences between the age and driving experience of participants between groups, $p > .05$. The visibility range was included as an environmental indicator for changes in automation uncertainty and varied between four equidistant levels ranging from low (91.44 m) to high visibility (1005.84 m)

(Met Office UK 2018). Thereby, variations in fog density were used to induce changes in visibility. This weather event was chosen for two reasons. First, the small water droplets that constitute fog scatter transmit pulses of sensors used for driving automation, leading to uncertainty increments (Rasshofer, Spies and Spies 2011). Second, fog was considered a salient environmental indicator for changes in automation uncertainty that is likely perceived and comprehended by novice users (Louw and Merat 2017).

In addition to the environmental indicator for changes in system uncertainty, one group, hereafter referred to as *experimental group*, was made aware of the system uncertainty through a dedicated display located in the instrument cluster. The other group, henceforth the *control group*, had no such instrument available. For the experimental group, the system uncertainty was conveyed using a stylised heartbeat whose rate was mapped to the visibility range and thereby to the uncertainty of the system (see Subsection 4.1.5). Mirroring human heart rate ranges, the upper end of the visibility range was mapped to a heart rate of 50 bpm, whereas the lower end was set to a maximum of 140 bpm.

To assess drivers' responses to emergency takeovers following system failures, the experiment implemented a TOR that prompted participants to manually control the vehicle following an extended period of automated driving. Contrasting closely related work (Beller, Heesen and Vollrath 2013), the study did not implement multiple system malfunctions. Instead, it was relied on a single failure (Molloy and Parasuraman 1996). While this reduces data available to analysis and thereby lessens statistical power, it allows for a more realistic failure rate. Implementing multiple takeovers in a laboratory scenario lasting less than an hour is not representative of real-world driving systems. While a single TOR still exceeds realistic failure rates, it allows for the conduction of empirical tests in a controlled environment without jeopardising participants. In addition to arguments relating to the realism of the scenario, HF related aspects further favour the use of one instead of multiple TORs. As such, complacent behaviour decreases and the detection of system failures deteriorates with longer periods of faultless automation performance (Davies and Parasuraman 1982; Parasuraman and Manzey 2010) (see Subsection 2.6.2.2). Thus, the strongest detrimental effects of extended system monitoring are likely observable with a single takeover.

In concordance with related work, the TOR was conveyed using a multimodal cue consisting of a visual alarm (instrument cluster flashing red at 3 Hz) and an auditory alarm (abstract pulses) (Bazilinskyy et al. 2018; Politis, Brewster and Pollick 2015b). The TOR was issued with 7 s TTC remaining (Gold, Damböck et al. 2013) and resulted in an immediate disengagement of the automated driving system. This contrasts related publications in which the TOR notified drivers about an imminent system deactivation (Large, Burnett, Salanitri et al. 2019). The immediate disengagement was implemented to account for a limitation of the driving simulation software that prevented user-initiated takeovers. Consequently, drivers had no means of deactivating the automated system prior to its planned disengagement and were therefore required to remain in the role of a passive monitor following the alarm – even if they already noticed a potential obstacle ahead and wanted to take over control. This behaviour is unlikely to be conformable with user expectations and was therefore avoided through an immediate system disengagement.

As outlined in Section 2.7 and highlighted in the integral model visualised in Figure 2.14, the presence of NDRTs affects operator behaviour. To account for this influence and to fulfil the initially defined requirement of a comprehensive evaluation, participants were engaged in an NDRT while the automation was activated. The NDRT was implemented as a visual search task, which is one of the most significant cognitive subtasks in the context of driving (see Subsection 3.2.3.4) (Shinohara et al. 2010). The visual search task prompted participants to identify if a predefined target is present among a collection of distractors (see Subsection 3.2.3.4). A positive response was required if either a blue letter or the letter S was visible (see Figure 3.5).

4.1.3 Dependent Variables

To assess the outlined research questions (see Section 4.1), measurements of the dependent variables trust, SA, gaze behaviour, workload, and operator heart rate were taken (see Subsection 3.2.3).

Trust was measured using single-item ratings on a percentage scale. This was preferred over more detailed trust scales (Jian, Bisantz and Drury 2000) to allow for the repeated, non-distracting measurement without interrupting the simulation scenario and interfering with participant behaviour (see Subsection 3.2.3.4). Specifically, the experimenter prompted the participant to answer the following question: 'To what percentage do you currently trust the automated system to safely perform the driving task?'

The SAGAT was applied to achieve an objective and direct measure of SA (Endsley and Garland 2000; Nguyen, Lim et al. 2019). Without prior notification of the participant, the simulation was stopped and all screens were blacked out. Using the following questions, the participant was immediately queried about the current driving situation:

1. On which lane are you?
2. Are there other vehicles around you?
3. What is the position of the vehicles around you?
4. What is the relative speed of the vehicle closest to you?

The questions were formulated to evaluate the participants' awareness of their surroundings, mainly related to the perceptional level of SA (see Subsection 2.6.2.2). Following the completion of the SAGAT questions, the simulation was continued from the point at which it was halted. To minimise the effect of the interruption on other dependent variables such as heart rate, the SAGAT was followed by a recovery section in which no measures of dependent variables were taken.

The monitoring behaviour of operators was measured using a combined direct (eye tracking) and indirect measurement (visual search task). In reference to related work (Hergeth et al. 2016), eye tracking data were analysed using monitoring frequency and ratio (see equations 3.3 and 3.4, Subsection 3.2.3.2). Within this experiment, eye tracking data was processed for a duration of 80 s prior to the TOR. This included two equally long sections with thick fog, i. e. low visibility and high uncertainty, and no/low fog, i. e. high visibility and low uncertainty. For a period between 80 s and 40 s before TOR, the visibility was high (TOR-80-40). The following 40 s up to the TOR were characterised by thick fog (TOR-40-0). A failure of the eye tracking storage unit prompted the exclusion of one participant for the assessment of monitoring frequency and ratio. The indirect measurement of operator gaze behaviour was realised using the solving rate of NDRTs. A larger number of solved tasks indicates an increased engagement in the NDRT. It was therefore expected that the solving rate of NDRTs and the monitoring frequency are inversely related. Takeover performance data was recorded following the TOR. Specifically, MTTC as well as lateral and longitudinal acceleration values were measured. Further, TTT was recorded as a measure for the time between TOR and initial driver input, manifesting itself in changes of steering angle, brake or accelerator position. Additionally, workload was measured using a subjective workload assessment questionnaire consisting of six 7-point Likert scale items analogous to the NASA-TLX (Hart 2006; Hart and Staveland 1988). Finally, the heart rate of participants was measured as an indicator for both stress and trust (Waytz, Heafner and Epley 2014).

4.1.4 Apparatus

The experiment was conducted in a soundproof laboratory with controlled lighting conditions. For the duration of all practice and experimental scenarios, participants were positioned in a driving simulator consisting of a mock-up Land Rover cockpit that faced three connected projection screens, each with an area of 160 cm × 90 cm (see Subsection 3.2.3.1). The projection screens were placed at a distance of 160 cm from the cockpit's steering wheel and generated a combined 135° field of view. The cockpit was equipped with a digital instrument cluster that displayed the uncertainty information. Further, an Apple iPad Air 2 was attached to the centre console and functioned as the interactive display for the visual search task. Tobii Pro 2 eye tracking glasses were used to record operator gaze behaviour at

Figure 4.2: Uncertainty display consisting of a stylised heartbeat and a numeric indication for beats per minute (Kunze 2018)

a sampling frequency of 100 Hz. The recorded eye tracking data was subsequently processed with the proprietary software Tobii Pro Lab (Tobii AB 2018b). The optical heart rate monitoring feature of a Polar A370 wearable device were used to measure the heart rate of participants with a frequency of 1 Hz.

4.1.5 Uncertainty Visualisation

The uncertainty was conveyed visually using a display positioned in the instrument cluster in order to allow for the comparability with related work that presented uncertainty content in a similar manner (Beller, Heesen and Vollrath 2013; Helldin et al. 2013).

To promote trust in automation and to convey uncertainties in a naturalistic fashion, an anthropomorphic visualisation of uncertainty was chosen (see Subsection 2.8.2.3). To address experimental research question 2 (see Section 4.1), the visualisation method had to be suitable for showing several levels of uncertainty. Beller, Heesen and Vollrath (2013) employed stylised facial expressions and gesticulations to indicate that the system is uncertain. However, the implementation was binary and did not allow for the display of uncertainties with a higher resolution. In fact, while facial expressions invoke the strongest anthropomorphism-related responses, subtle changes in facial expressions that are needed to implement several levels demand large attentional resources and are difficult to identify (Gollan et al. 2010). Thus, alternative human-like features should be contemplated.

Beller, Heesen and Vollrath (2013) interpreted uncertainty as a state of ambiguity, thus mapping the system state directly to a visual equivalent. Alternatively, downstream responses to uncertainty can be considered. As such, uncertainty requires additional cognitive processing to cope with the ambiguity of the situation and likely induces stress. When humans experience stressful situations – for instance taking a critical decision in an uncertain situation – a series of physiological symptoms indicate the increase in stress levels, among which is an increase in heart rate (Hubbard and Workman 1998; Steptoe and Voegele 1991). Additionally, human heart rate was shown to raise in situations that require higher levels of focus and alertness (Wang, Pinol et al. 2014). It is therefore proposed that the visualised simulated heart rate of an automated system may function as an alternative anthropomorphic indication for uncertainty.

Within this experiment, the uncertainty information was presented to participants of the experimental group with a stylised, animated heartbeat accompanied by a decimal signifying beats per minute (see Figure 4.2). Increases in heart rate were indicated by showing a red colour on the whole instrument cluster for a duration of 500 ms to accommodate the need for stimuli that attract bottom-up attention (see Subsection 2.6.2.2).

4.1.6 Procedure and Scenario

Pre-recorded videos were used to explain the objectives, framing, and procedure of the experiment. The following is an excerpt of the information provided to participants:

> Driving automation is fast approaching. This will be evolutionary rather than revolutionary. This means we will not have an immediate transition from manual driving to automated driving, but rather a gradual process. For the foreseeable future, human operators will need to take over the driving task if the automation reaches its limits. Vehicle automation is still far from perfect. There are many situations in which automated vehicles are experiencing difficulties.

Following the video-based briefing, participants were asked to fill in a consent form and to report demographic as well as driving-related personal data.

While seated in the driving simulator, participants received an instruction regarding the properties of the simulated vehicle. Specifically, participants were informed that the automation system was capable of performing both the lateral and longitudinal component of the driving task on a sustained basis and were made aware of the consequences of potential system malfunctions, i. e. TORs. Participants were instructed to immediately take over the complete driving task following a TOR. Members of the experimental group were additionally introduced to the uncertainty display. Differences in instructions were shown to result in effects on dependent variables, for instance trust (Körber, Baseler and Bengler 2018). Therefore, all instructions were scripted to describe the system in a consistent manner.

Succeeding the instructions, participants completed a series of practice scenarios with a combined duration of 30 to 40 minutes. Each scenario consisted of a drive on a two-lane highway and incorporated several emergency evasive and/or braking manoeuvres. Moreover,

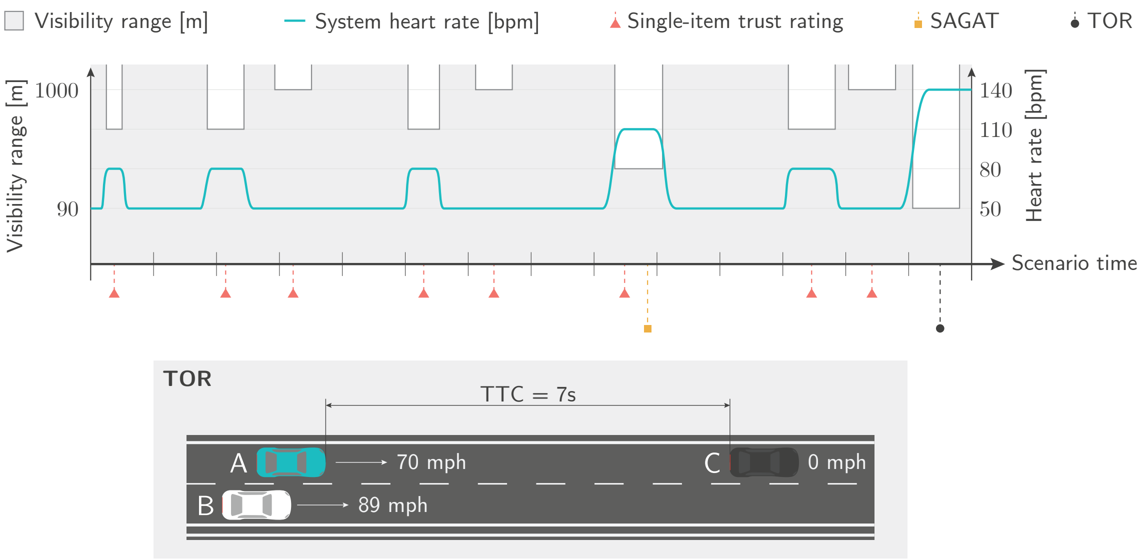


Figure 4.3: Visibility range and system heart rate across the experimental scenario (top); location of own vehicle (A), passing vehicle (B), and stopped vehicle (C) at the time of the TOR (bottom)

participants were presented with a TOR. After completing the practice scenario, participants were equipped with the wearable device for heart rate measurement and the eye tracking glasses, for which a calibration was performed (see Subsection 3.2.3.2). Seated in the simulator, participants were asked to complete a series of practice visual search tasks.

Following a short rebriefing about the experiment and a final opportunity to ask questions, the experimental scenario commenced. The scenario lasted for a total duration of 20 minutes and was situated on a two-lane UK motorway with moderate traffic. To replicate scenarios in which an automated driving system would be activated upon entering a motorway, the automation was initially disengaged and participants were instructed by a prerecorded voice to continue driving manually until an overhead gantry was crossed. Immediately following the overhead gantry, the automated driving system was engaged and participants were instructed to centre the steering wheel. The correct centring of the steering wheel was controlled by the experimenter in order to avoid inadvertent manoeuvres at the point of automation reactivation. The simulated vehicle, implemented as an SAE level 3 driving automation system, kept its speed constant at 70 mph and performed a series of lane change and overtaking manoeuvres. Figure 4.3 illustrates how the visibility rate fluctuated through the scenarios and indicates when measures of SA and trust were taken and the TOR was issued.

The system failure prompting a TOR was modelled in reference to recent crashes of vehicles equipped with automated driving systems. As such, a vehicle had stopped in the lane of the ego vehicle, leading to an uncertain condition of the automated driving system (Shepardson 2018; Stewart 2018). Further, another vehicle passed the ego vehicle on the right hand lane at the time of the TOR (see Figure 4.3). The takeover scenario was designed to require SA on all three levels (see Subsection 2.6.2.2). As such, participants had to notice both the overtaking and the stopped vehicle, understand that a lane change was not immediately possible, and predict the overtaking vehicle's position in order to plan a future manoeuvre.

The SAGAT test was conducted in a scenario that replicated the takeover situation, with the exception of the stopped vehicle. This afforded the measurement of SA in a situation comparable to that of the takeover without interrupting it.

Following the completion of the experimental scenario, participants filled in the workload questionnaire. The experiment was concluded with a short unstructured interview, after which the participants were debriefed.

4.1.7 Data Analysis

For each participant, multiple measurements of the variables solving rate of NDRTs, trust, and monitoring frequency were taken. While this increases validity by reducing the risk that the measurements are due to chance, it results in several non-independent data points, thereby prompting the use of LMMs (see Subsection 3.2.5). Unless explicitly stated otherwise, random effects were implemented for subjects using intercepts. To account for the ordinal characteristics of the independent variable visibility range, it was treated as continuous. The presence of the uncertainty display was, in contrast, added to the model as a categorical variable. Residual plots were visually inspected to validate the assumption of homoscedasticity. The R package lmerTest (Kuznetsova, Brockhoff and Christensen 2018) was applied to approximate degrees of freedom and p values for t statistics based on Satterthwaite's method.

Outliers regarding driving performance data that were not immediately linked to the takeover manoeuvre were removed to prevent erratic effects on results. In one instance, the maximum lateral acceleration was recorded 15 s after the TOR. The corresponding video sequence and vehicle position metrics clarified that the ego vehicle had, at that point, already passed the stopped vehicle and that the manoeuvre which led to the increased acceleration value was not connected to the evasive manoeuvre following the TOR.

4.2 Results

In the following, the results of the performed statistical analyses are presented relative to the corresponding experimental research questions (see Section 4.1 and Table 1.1).

Table 4.1: Results of the performed linear mixed effects models with approximated degrees of freedom and p values based on Satterthwaite's method

Variables	*Estimate*	*SE*	*df*	*t*	*p*
Solving rate of NDRTs					
Visibility range	0.368	0.460	304.000	0.800	.424
Uncertainty display	−15.007	2.396	91.350	−6.264	<.001***
Interaction	3.196	0.650	304.000	4.914	<.001***
Solving rate of NDRTs					
Subjective trust	0.064	0.023	268.070	2.770	.006*
Uncertainty display	−6.874	1.704	32.360	−4.033	<.001***
Subjective trust					
Visibility range	0.605	1.591	236.000	0.380	.704
Uncertainty display	−48.052	7.484	147.500	−6.420	<.001***
Interaction	15.286	2.250	236.000	6.795	<.001***
Monitoring frequency					
Uncertainty display	0.385	0.056	44.160	6.877	<.001***
Section	0.004	0.032	32.000	0.137	.892
Interaction	−0.216	0.046	32.000	−4.752	<.001***
Subjective trust					
Monitoring frequency	−29.184	13.529	12.310	−2.157	.039*
Uncertainty display	3.288	4.803	31.000	0.685	.499
Solving rate of NDRTs					
Monitoring frequency	−22.385	5.798	62.950	−3.861	<.001***
Uncertainty display	−1.648	3.079	38.680	−0.535	.596

*$p < .050$ **$p < .005$ ***$p < .001$

4.2.1 Impact on Subjective Trust, SA and Monitoring Behaviour

4.2.1.1 Subjective Trust

The impact of dynamically conveying uncertainties on subjective trust was evaluated using an LMM, whereby the independent variables uncertainty display and visibility range as well as their interaction were implemented as fixed effects (see Table 4.1). The interaction between the independent variables was significant, $t(236.000) = 6.795, p < .001$. Specifically, the visibility range only predicted subjective trust ratings when the uncertainty display was present (see Figure 4.4). In addition, the model returned a main effect of the uncertainty display. As visualised in Figure 4.4, the subjective trust ratings were not consistently lower

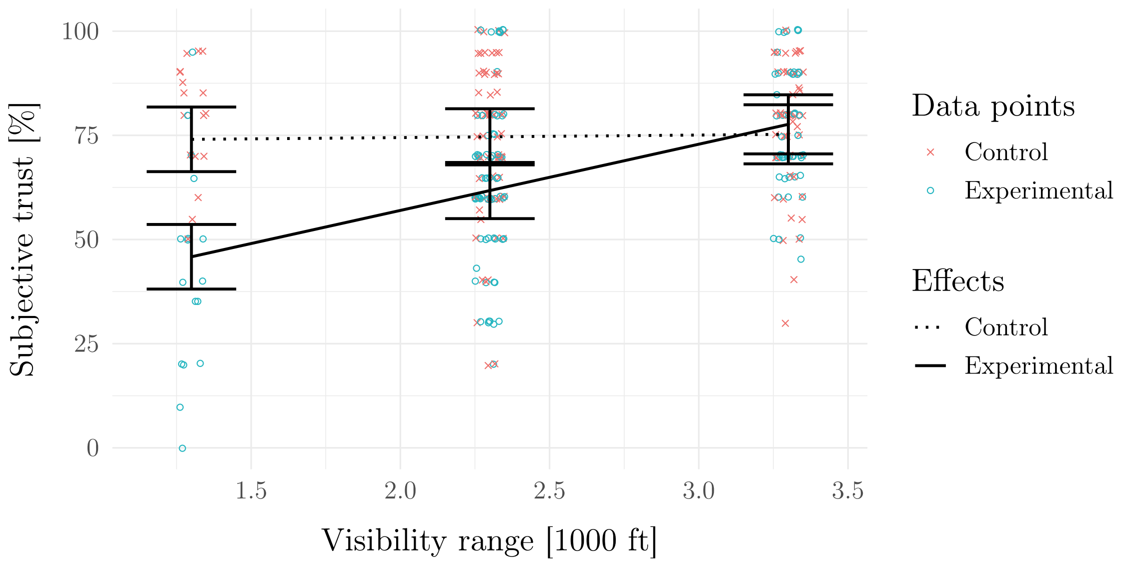


Figure 4.4: Effects of the conducted linear mixed effects model predicting the impact of visibility range and uncertainty display on subjective trust. Jitter with a width of 0.05 was added to data points to improve visibility.

for the experimental group. Rather, the subjective trust ratings were – as indicated by the significant interaction – only lower with decreasing visibility (see Table B.1). Thus, the significant main effect of conveying uncertainties is the consequence of significant variations due to the interaction between the independent variables and can therefore not be considered meaningful. Hence, the sole presence of the uncertainty display did not show an impact on trust. Instead, changes in the conveyed information affected subjective trust ratings.

4.2.1.2 SAGAT Responses

The ratios of correctly answered SAGAT questions by group are summarised in Table 4.2. For each of the SAGAT questions (see Subsection 4.1.3), participants of the experimental group provided, on average, more correct answers. A chi-squared test with Yates' continuity

Table 4.2: Ratio of correct responses for each SAGAT item. For instance, the ratio of 0.471 in the leftmost data cell of the first row indicates that 47.1 % of participants in the control group correctly responded to SAGAT item 1 (see Subsection 4.1.3).

Group	**SAGAT 1**	**SAGAT 2**	**SAGAT 3**	**SAGAT 4**
Control	0.471	0.588	0.412	0.471
Experimental	0.765	0.882	0.529	0.529

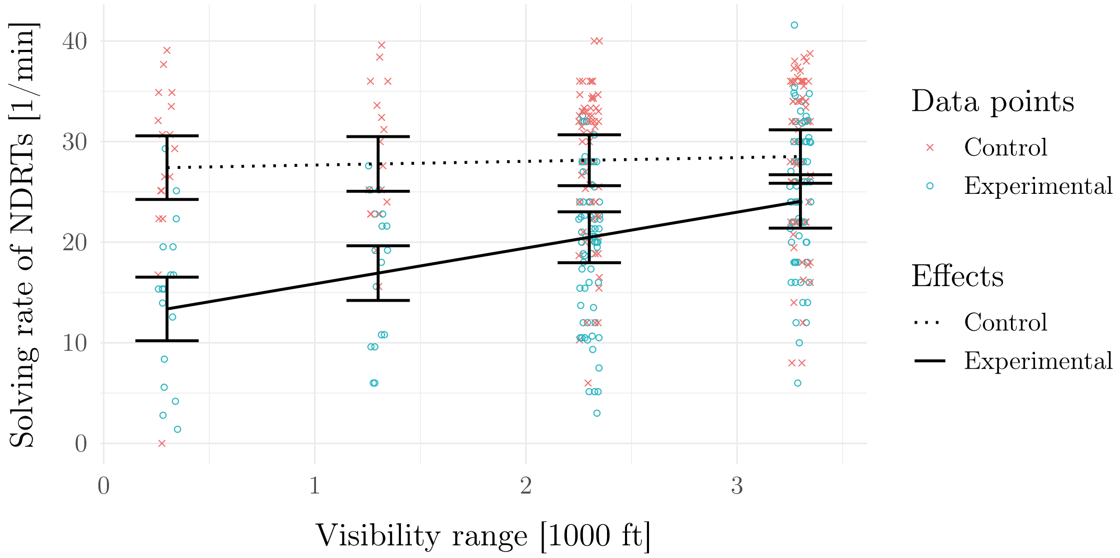


Figure 4.5: Effects of the conducted linear mixed effects model predicting the impact of visibility range and uncertainty display on the solving rate of NDRTs. Jitter with a width of 0.05 was added to data points to improve visibility.

correction was performed to assess the differences between the ratios. The difference between ratios was significant, $\chi^2(1) = 4.349, p = .037$.

4.2.1.3 Monitoring Behaviour

The monitoring behaviour of operators was evaluated using the solving rate of the visual search task and the recorded eye tracking data.

The impact of dynamically conveying uncertainties on the solving rate of the visual search task was evaluated using an LMM, whereby the independent variables uncertainty display and visibility range as well as their interaction were implemented as fixed effects. The model returned a significant interaction, indicating that the visibility range predicted the solving rate dependent on the presence of the uncertainty display, $t(304.000) = 4.914, p < .001$. Figure 4.5 highlights that the visibility range only predicted the solving rate if the uncertainty display was present. Further, the mean solving rate for participants of the experimental group was consistently lower than that of the control group (see Table B.2). Thus, the presence of the display may have led to a reduction in solved NDRTs.

The monitoring frequency to driving-related AOIs was used to further assess the impact of the independent variables on monitoring behaviour. An LMM was performed on monitoring frequency with the fixed factors uncertainty display and section (TOR-80-40, TOR-40-0, see Subsection 4.1.3) as well as their interaction. The model returned a significant main

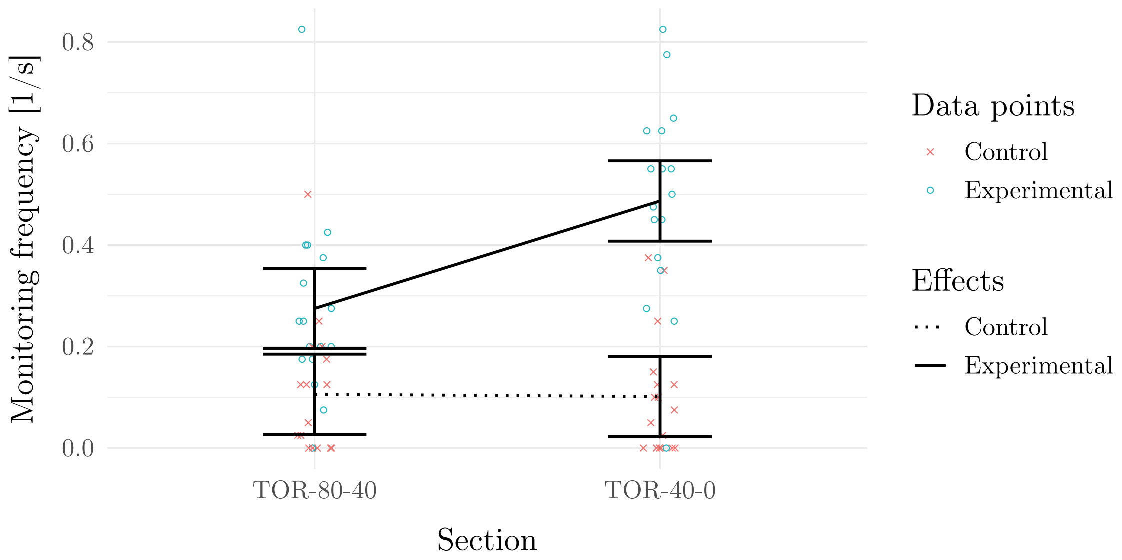


Figure 4.6: Monitoring frequency towards driving-related AOIs in a low uncertainty section (TOR-80-40) and high uncertainty section (TOR-40-0)

effect of the uncertainty display, $t(44.160) = 6.877, p < .001$, and a significant interaction, $t(32.000) = -4.752, p < .001$. The effects of the LMM are depicted in Figure 4.6. The monitoring frequency only increased between sections for the experimental group. Further, the control group had, on average, a lower monitoring frequency for both sections (see Table B.3), confirming the main effect of the uncertainty display.

The monitoring ratio was evaluated as a third indicator for monitoring behaviour. The results are summarised in Figure 4.7 and Table B.4. Similar to the recorded monitoring frequency, the control group did not adjust their monitoring behaviour between the two sections. Contrasting this, participants of the experimental group focused more on the FRD ($MD = 0.219$), uncertainty display ($MD = 0.073$), and the instruments ($MD = 0.009$) while allocating less attention on the NDRT in the second, high uncertainty section (TOR-40-0).

4.2.2 Monitoring Behaviour and Subjective Trust

Previous research has indicated that operator trust in automation capabilities predicts monitoring behaviour (Hergeth et al. 2016). Being an essential component of the integral model presented in Figure 2.14, the relationship between the two variables was assessed to potentially reaffirm the results.

Figure 4.8 visualises the solving rate of NDRTs as an indirect indicator of monitoring behaviour relative to subjective trust ratings. A corresponding LMM returned significant

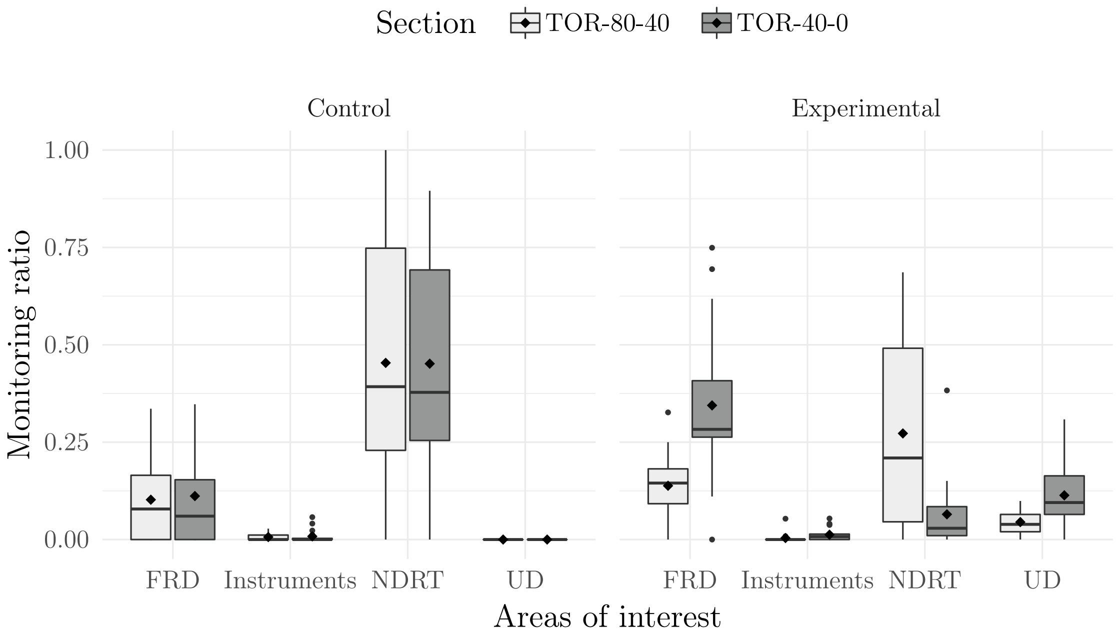


Figure 4.7: Monitoring ratio for the AOIs FRD, instruments, NDRT, and uncertainty display (UD) regarding two sections between groups

results, $t(268.070) = 2.770, p = .006$, indicating that subjective trust significantly predicts the solving rate of NDRTs and indirectly monitoring behaviour.

An additional LMM was performed using monitoring frequency as a direct indicator for monitoring behaviour (see Figure 4.9). Again, a significant main effect was identified, $t(12.310) = -2.157, p = .039$.

The significant effects of both indirect and direct measurements for monitoring behaviour suggest that they predict each other. In fact, a further LMM confirmed this assumption, $t(62.950) = -3.861, p < .001$.

4.2.3 Impact of Higher Resolution

Tables B.1 and B.2 summarise the descriptive statistics for subjective trust indications and NDRT solving rate depending on visibility range and group. Regarding the experimental group, the collected data indicates a gradual change in both subjective trust and NDRT solving rate with increasing visibility range. For instance, subjective trust increased from 42.941 % ($SD = 25.190$ %) at a visibility range of 1300 ft via 63.206 % ($SD = 19.400$ %) at a visibility range of 2300 ft to 76.667 % ($SD = 14.549$ %) at a visibility range of 3300 ft. A similar trend could not be observed for the control group.

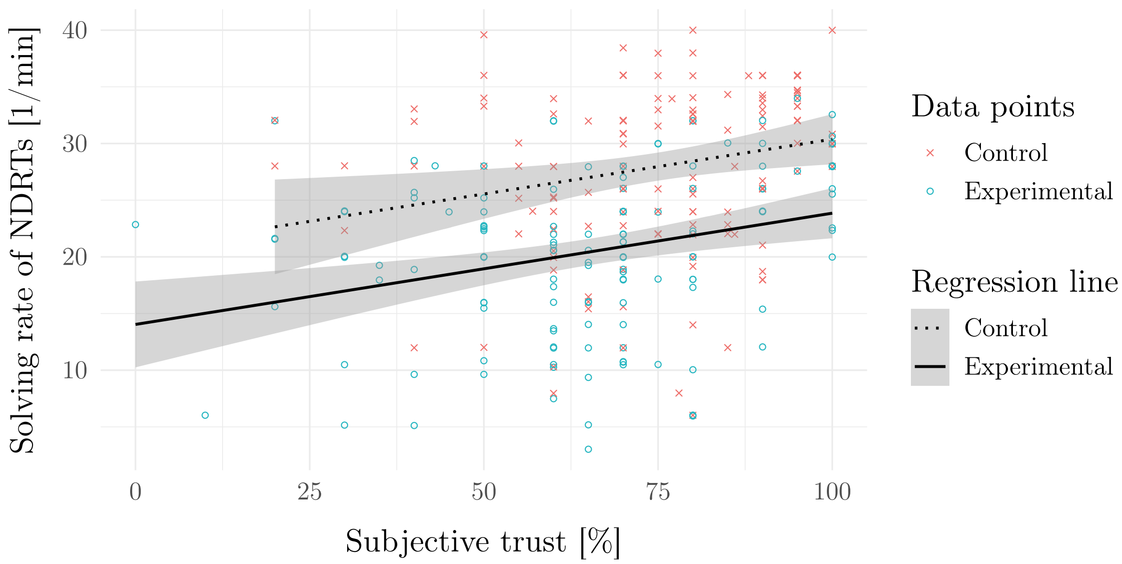


Figure 4.8: Solving rate of NDRTs relative to subjective trust and uncertainty display

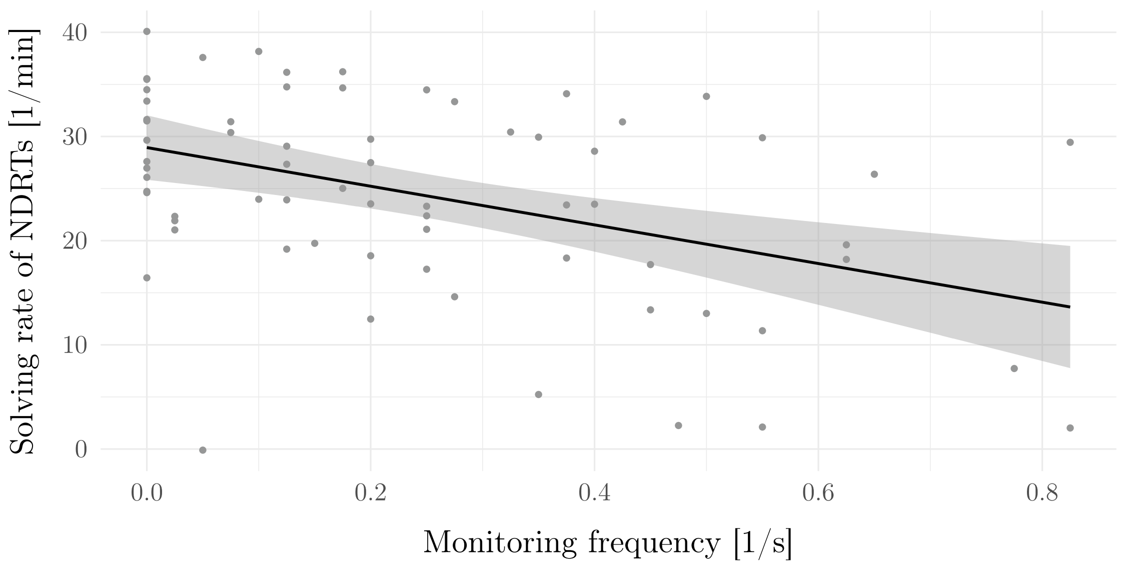


Figure 4.9: Monitoring frequency relative to solving rate of NDRTs

4.2.4 Takeover Performance

4.2.4.1 Minimum Time to Collision

Significantly higher MTTC values were recorded for participants of the experimental group ($M = 2.765\,\mathrm{s}$, $SD = 1.661\,\mathrm{s}$) compared with the control group ($M = 1.641\,\mathrm{s}$, $SD = 1.217\,\mathrm{s}$), $t(29.329) = -2.251, p = .032, d = 0.772$ (medium effect size).

4.2.4.2 Maximum Lateral and Longitudinal Acceleration

Vehicle dynamics data, particularly lateral and longitudinal acceleration, are considered indicators for driving quality and can be evaluated to assess the quality of takeovers (Bubb et al. 2015). Maximum lateral acceleration values following the TOR did not significantly differ between the experimental group ($M = 2.021\,\mathrm{m\,s^{-2}}$, $SD = 1.046\,\mathrm{m\,s^{-2}}$) and the control group ($M = 2.278\,\mathrm{m\,s^{-2}}$, $SD = 1.232\,\mathrm{m\,s^{-2}}$). There was also no significant difference in maximum longitudinal acceleration values between groups.

4.2.4.3 Time to Takeover

TTT values were, on average, higher for the control group ($M = 2.356\,\mathrm{s}$, $SD = 1.604\,\mathrm{s}$) than for the experimental group ($M = 2.090\,\mathrm{s}$, $SD = 1.614\,\mathrm{s}$). The differences were, however, not significant at the 5 % level.

4.2.5 Workload

The responses to the NASA-TLX workload questionnaire are summarised in Table 4.3. Two-sample (unpaired) t tests were performed to assess the differences between groups for each questionnaire item. Participants reported experiencing significantly greater perceived effort when using the uncertainty display ($M = 5.529$, $SD = 0.800$) compared without the display ($M = 3.706$, $SD = 1.532$), $t(24.123) = -4.352, p < .001$. The remaining items were not significant at the 5 % level.

4.2.6 Physiological Measures

The heart rate of participants was measured as a potential indicator for stress and trust. Neither the presence of the uncertainty display nor the visibility range or trust ratings significantly predicted the absolute or normalised heart rate of participants ($p > .05$). In line with prior work (Waytz, Heafner and Epley 2014), the heart rate did, however, change significantly as a consequence of the TOR, $t(1359) = -7.840, p < .001$.

Table 4.3: Descriptive statistics for NASA-TLX workload subscales

Workload subscales	**Statistic**	**Control**	**Experimental**
Mental demand	M	4.647	4.882
	SD	1.656	1.409
Physical demand	M	2.235	2.471
	SD	1.480	1.419
Temporal demand	M	4.588	5.118
	SD	1.543	1.409
Performance	M	5.294	5.000
	SD	0.985	1.620
Effort	M	3.706***	5.529***
	SD	1.532	0.800
Frustration	M	3.235	3.294
	SD	2.223	1.829

$^{*}p < .050$ $^{**}p < .005$ $^{***}p < .001$

4.3 Discussion

The previously described experiment was conducted to systematically evaluate the implications of dynamic uncertainty communication. As such, effects on trust, SA, monitoring behaviour, takeover performance, workload, and physiological measures were examined.

In line with other work from within the automotive domain (Beller, Heesen and Vollrath 2013; Helldin et al. 2013) and other research areas (McGuirl and Sarter 2003), the outcomes suggest a series of ramifications of dynamic uncertainty communication on human-automation interaction.

4.3.1 Impact on Subjective Trust, SA and Monitoring Behaviour

The results suggest that the availability of real-time uncertainty information has measurable effects on all dependent variables in question, namely subjective trust, SA, and monitoring behaviour.

For the experimental group, the visibility range significantly predicted subjective trust responses. Contrasting this, the trust of the control group did not change with the visibility range. This suggests that the uncertainty display supported users in assessing the current system reliability, leading to more appropriate trust.

As predicted, the awareness of the system's uncertainty – and thereby fallibility – prompted participants to adjust their behaviour. Specifically, the experimental group solved significantly less visual search tasks with increasing uncertainty, i. e. with decreasing visibility range. In contrast, the control group showed no such behavioural adaptation. This outcome supports the previously anticipated impact of uncertainty communication on endogenous operator attention allocation, as depicted in Figure 2.14. Both the differences in monitoring frequency and monitoring ratio between groups reaffirm the attentional shift from non-driving-related to driving-related AOIs with increasing uncertainty.

These results highlight that participants of the experimental group used the uncertainty information to monitor system behaviour and surroundings in situations that were likely to invoke a TOR. It was hypothesised that this change in attention allocation would result in measurable benefits to SA. In fact, the responses to SAGAT questions confirmed that participants of the experimental group were more aware of their surroundings in situations with increased uncertainty.

Nonetheless, the communication of uncertainties was not without side effects. Specifically, the solving rate of NDRTs was significantly lower for the experimental group, even for sections when the uncertainty was minimal. Both monitoring frequency and ratio indicate that participants of the experimental group focused more on the FRD and the uncertainty display, irrespective of the visibility range. This highlights a major drawback of visually presenting uncertainty information in the instrument cluster. While the control group was able to focus their attention on the NDRT in low uncertainty situations, the uncertainty display prompted the experimental group to regularly change their focus and monitor the uncertainty information presented in the instrument cluster to perceive potential changes. Even if operators employ an optimal attention allocation strategy, changes may be missed (see Subsection 2.6.2.2) (Moray and Inagaki 2000), particularly with highly reliable aids such as automated driving systems. Aside from the detrimental effects on NDRT performance, the visual instrument cluster display may also affect safety. Immediately before the TOR, participants of the experimental group monitored the uncertainty display with a ratio of 12.06 %, i. e. 4.82 s out of 40 s. Ideally, this share should be minimised in order to reduce the glance time off-road, which was shown to increase crash risk (Seppelt and Lee 2007). One approach to minimise this ratio is to employ an improved display design that allows drivers to peripherally perceive the uncertainty information (see Chapters 5 to 7). The need for displays that do not rely on focal visual attention is reaffirmed by Large, Burnett, Morris, Muthumani and Matthias (2017) who found that the prolonged use of such a display leads to missed changes.

In contrast to expectations and previous work (Beller, Heesen and Vollrath 2013), the anthropomorphic uncertainty display did not lead to higher trust scores compared with the control group in situations for which the uncertainty was minimal. This may be attributed to several reasons. First, the general awareness of the system's imperfections may lead operators to report more conservative trust values. Second, the anthropomorphic feature, heart rate, may not be sufficient to invoke effects that can be attributed to anthropomorphism. Rather, more easily recognisable features such as faces may be necessary (Visser, Monfort et al. 2017).

Overall, the discussed findings reaffirm the interrelations presented in the integral model (see Figure 2.14). Specifically, knowledge of system uncertainties allows operators to calibrate their trust and, consequentially, adjust their (endogenous) attention allocation strategy, culminating in an improved SA prior to takeovers.

4.3.2 Monitoring Behaviour and Subjective Trust

In line with previous publications (Hergeth et al. 2016; Muir 1987), the results of this experiment confirm the inversely proportional relationship between monitoring behaviour and subjective trust. As such, increases in trust correlate with decreases in attention allocation to driving-related AOIs.

4.3.3 Resolution

Consistent with the uncertainty display proposed by Helldin et al. (2013) but in contrast to the binary uncertainty indication applied by Beller, Heesen and Vollrath (2013), the uncertainty was conveyed using multiple levels. It remains to be addressed how the higher informational resolution affected human-automation interaction.

As depicted in Figures 4.4 and 4.5 and summarised in Tables B.1 and B.2, participants of the experimental group adjusted their trust in agreement with the individual uncertainty levels and allocated their attention to the NDRT accordingly. This shows that participants were not only able to distinguish distinct uncertainty levels but also to gauge their responses according to the perceived changes.

Thus, the focus on non-driving-related AOIs decreased from low via intermediate to high uncertainty levels. Whereas a binary communication of uncertainties would prompt drivers to focus all their attention on the driving scene whenever the uncertainty was higher than a prespecified threshould, a more gradual display allows drivers to gauge their responses and thereby allows for an optimised attention allocation. This finding is in line with Bisantz, Marsiglio and Munch (2005) who discovered that an increase in detail of uncertainty information leads to performance enhancements.

4.3.4 Takeover Performance

The driving performance data confirms that the communication of uncertainties leads to improvements in takeover performance and quality. In line with prior work (Beller, Heesen and Vollrath 2013; Helldin et al. 2013), the availability of an uncertainty display led to significantly higher MTTC values, indicating that the previously identified benefits to SA had positive implications for takeover performance. The recorded maximum lateral acceleration values by far exceeded the limits for comfortable driving (see Subsection 3.2.3.1). Consequently, participants of both groups performed rather abrupt manoeuvres.

Whereas participants of the control group likely had to build up SA from the moment of the TOR, participants of the experimental group were able to build up SA before the TOR was issued, as indicated by the previously described monitoring data. Since basic levels of SA were present at the time of the TOR, the experimental group required less time to plan and execute a manoeuvre, which is in line with the described driving performance data.

4.3.5 Workload

The responses to the NASA-TLX workload questionnaire indicate that both mental and temporal demand exceeded optimal levels, which are considered to be at around half of the scale (Stanton, Dunoyer and Leatherland 2011). Post-experiment interviews suggest that participants reported a high mental workload due to the demands of the divided attention setting that required the simultaneous handling of a NDRT and the monitoring of the FRD. Temporal demand was considered high due to the TOR, which was mostly perceived as abrupt and required participants to quickly manoeuvre the vehicle.

The groups differed in terms of perceived effort, whereby the experimental group reported higher levels. Interview responses suggest that the monitoring of an additional instrument, namely the uncertainty display, caused the high effort values. The participant statements are confirmed by eye tracking data, which indicates that participants regularly rotated their attentional focus around the AOIs NDRT, FRD, and uncertainty display (see Subsection 4.2.1). As described in Figure 2.7, movements in the eye or head field – which is required for the described change in gaze – result in higher information access effort.

Particularly for extended usage, the higher information access effort may lead users to ignore the additional content presented in the uncertainty display, especially after prolonged periods of safe automation performance (Large, Burnett, Morris, Muthumani and Matthias 2017). Consequently, a display that allows users to perceive the the system uncertainties without an increase in information access effort is needed.

4.3.6 Physiological Measures

The heart rate of participants was measured as an indicator for stress and trust. Particularly the implementation of the uncertainty display using a machine heartbeat might have influenced that of participants. However, the recorded absolute and normalised heart rate data does not confirm this. Further, heart rate did not predict trust and did not differ significantly between groups, but increased significantly as a consequence of the TOR.

4.4 Limitations

The discussed results of this experiment are delimited by the following limitations. Consistent with prior work (Beller, Heesen and Vollrath 2013; Helldin et al. 2013; Hergeth et al. 2016), the driving scenario was identical for all participants. As a consequence, the results may be limited to the specific fluctuations in fog density selected for testing. This trade-off was deliberately accepted to ensure that the scenario evokes maximum complacency effects for all participants, which was only feasible by placing the TOR near the end of the scenario (Lee and Moray 1992, 1994; Parasuraman and Manzey 2010).

Additionally, changes in visibility range as a consequence of fog density were used as an environmental indicator for automation uncertainty. The results indicate that this led to no behavioural adjustment of the control group. Other factors such as traffic density, missing lane markings, or weather conditions other than fog may have a different impact on operator behaviour and trust. It can also be argued that the findings do not have external validity as fog would be perceived differently in a real-world scenario, i. e. not just on a two-dimensional plane but rather three-dimensionally.

Further, the impact of uncertainty communication on physiological measures was solely based on recorded heart rate data. Other indicators such as HRV may lead to differing results. However, related work has suggested that HRV is not a significant predictor for trust (Wintersberger, Sawitzky et al. 2017).

4.5 Summary and Conclusions

The empirical results obtained from the conducted experiment confirm the hypothesised interrelations between trust, attention allocation, SA, and takeover performance that were previously outlined in the integral model (see Subsection 2.7.2 and Figure 2.14).

In agreement with steps 1 and 2 of the integral model (Beller, Heesen and Vollrath 2013; Hoff and Bashir 2015), uncertainty communication affected subjective operator trust indications and, consequently, led to more appropriate trust, as indicated by the significant prediction of trust by system uncertainty (or visibility range). Corresponding to steps 3 and 4, this affected the top-down attention allocation of operators (Hergeth et al. 2016; Muir and Moray 1996). Specifically, the monitoring frequency and ratio of driving-related AOIs increased while the solving rate of NDRTs decreased with lower visibility range, i. e. higher uncertainty. Thereby, participants showed a gradual response to changes in uncertainty. As such, monitoring behaviour and NDRT performance differed between low, intermediate, and high uncertainties, indicating that participants are able to accurately gauge their responses according to the actual demand. Closing the response chain depicted in Figure 2.14, SAGAT ratings and takeover performance data indicate higher SA, thereby affirming step 5. Thus, through a series of cognitive and behavioural responses, uncertainty communication was shown to positively affect operator SA prior to and during critical situations.

Nonetheless, the chosen implementation did entail detrimental implications to human-automation interaction. As such, the communication of uncertainties via the instrument cluster induced an increased perceived effort relative to the control group due to the necessity of regularly monitoring the display for changes. This further led to overall lower NDRT performance compared with the control group and resulted in increments to the glance time off-road. Consequently, an uncertainty communication method should be developed that exploits the described benefits while reducing the information access effort and the chances of missed signals.

CHAPTER 5

Designing for Peripheral Uncertainty Communication

Abstract The use of a visual uncertainty display positioned in the instrument cluster requires users to regularly change their attentional focus to perceive changes, leading to several HF issues that derogate the benefits of uncertainty communication (see Chapter 4). As such, the instrument cluster display leads to an increased information access effort, a higher likelihood of missed changes, and NDRT performance decrements. To overcome the previously identified issues, the information must be conveyed in a way that affords users the perception of uncertainties without the need for changes in gaze. By definition, peripheral awareness displays are suited for fulfilling this requirement. As a prerequisite for the evaluation of such a display for uncertainty communication, it must first be ensured that the developed display indeed affords users to be peripherally aware of information. Existing research, however, lacks a comprehensive framework and process for the development of such displays. Addressing this shortcoming, the current chapter presents a novel end-to-end approach to the development of peripheral awareness displays based on a consolidation of existing research. In addition to publications that are directly related to peripheral awareness displays, relevant ISO norms and user experience models are taken into account to ensure a holistic perspective. The consolidated development process serves as a reference for the design of a peripheral awareness display for uncertainty communication in the subsequent chapter. Thereafter, the resulting display is compared with the previously evaluated instrument cluster display to identify the implications of peripheral uncertainty communication and assess if the HF issues were alleviated. Aspects of this chapter, in part verbatim, are published in Kunze, Summerskill et al. (2019b).

Background	Objectives	Method	Results	Implications
Using a digital instrument cluster to convey uncertainties requires operators to regularly change their gaze to perceive changes, leading to increased workload.	Investigate peripheral awareness displays as a means for conveying information and consolidate the findings to a holistic evaluation method and framework.	Systematic literature review of frameworks and processes that support the design, development, and/or evaluation of peripheral awareness displays.	Comprehensive framework for the development and evaluation of peripheral awareness displays that considers pragmatic, hedonic, and emotional aspects.	The consolidated evaluation framework and process are subsequently applied to design a peripheral uncertainty display that reduces workload increments.

Figure 5.1: Overview of Chapter 5

5.1 Introduction

The majority of information that we perceive is not at the centre of our attention, it rather lies in its periphery (Matthews, Dey et al. 2004). On a daily basis we monitor objects and processes in the environment without consciously realising it – unless an event or signal causes a switch in attention. Peripheral awareness displays make use of this ability to perceive information without the need for focal attention (see Section 2.4).

Numerous terms describing displays that operate in the periphery of the user exist in the literature, notably *peripheral*, *notification*, and *ambient* displays (Matthews, Rattenbury and Carter 2007; Matthews, Dey et al. 2004; McCrickard, Chewar et al. 2003; McCrickard and Chewar 2003; Mankoff, Dey et al. 2003; Shami, Lshed and Klein 2005; Pousman and Stasko 2006; Wisneski et al. 1998; Stasko et al. 2004). Consulting previous publications, common themes among these terms can be discussed despite the absence of clear-cut definitions.

The term *peripheral displays* is often used as an umbrella term for interfaces that (a) are not within the user's primary focus of attention (see Subsection 2.4.3) and (b) function as a tool for communicating supporting information regarding at least one of the user's secondary activities (Stasko et al. 2004; Matthews, Rattenbury and Carter 2007; Matthews, Dey et al. 2004). Thus, *peripheral displays* afford users to attend task A with focused attention while keeping them aware of information regarding task B. As a consequence, users can balance several tasks without being overburdened (Matthews, Rattenbury and Carter 2007; Weiser and Brown 1996).

McCrickard, Chewar et al. (2003, p. 315) define *notification systems* 'as interfaces that are typically used in a divided-attention, multitasking situation, attempting to deliver current, valued information through a variety of platforms and modes in an efficient and effective manner'. Further, McCrickard and Chewar (2003, p. 67) describe *notification displays* as interfaces that provide access to 'information from sources secondary to current activities'. As such, the terms *peripheral* and *notification* displays or systems can be used almost interchangeably.

Matthews, Rattenbury and Carter (2007, p. 234) argue that notification displays focus on *important information*. However, this is based on a misinterpreted citation. The actual statement that Matthews, Rattenbury and Carter (2007) refer to does not contain an emphasis on the importance of information, thus removing the basis of their argument to distinguish between notification systems and their suggested term, peripheral displays. Their proposed discrimination between the two terms can further be denounced since McCrickard, Chewar et al. (2003, p. 328) apply their proposed model to low criticality applications, such as informative art.

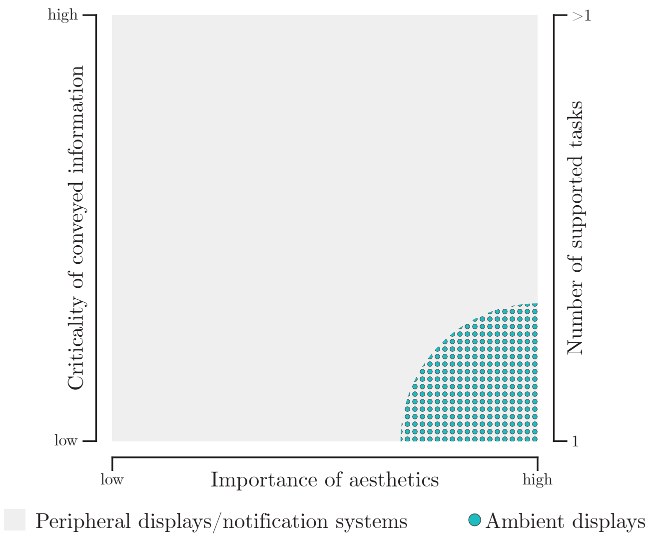


Figure 5.2: Peripheral display types

Ambient displays can be considered a proper subset of both peripheral and notification displays, focusing on the aesthetic and often abstract communication of important but not critical information regarding mostly a single secondary activity (see Figure 5.2) (Pousman and Stasko 2006; Mankoff, Dey et al. 2003; Stasko et al. 2004). Thereby, changes in perceptual cues are meant to be subtle and often indicate the state of larger systems (Wisneski et al. 1998).

Common among all described definitions is the emphasis on the awareness of the existence of another information source in the periphery of attention. While McCrickard, Chewar et al. (2003) were first to describe a framework for suchlike systems, the term *notification* implies a high formality of the conveyed information (Oxford Dictionaries 2018c), which itself suggests a high importance of that information, corresponding to a likely interruption. However, suchlike systems should only interrupt the user when the information is critical. Using only the term *peripheral*, on the other hand, does not specifically consider the attentional and perceptional aspects. Thus, a *peripheral display* could be any display that is not currently in the centre of our visual focus, irrespective of whether it is perceived or not. Hence, the term *peripheral awareness* will be used to emphasise on the defining features of suchlike tools, i. e. the awareness of a display that conveys information regarding one of the user's tasks without being at the centre of attention but rather in its periphery.

Further, it needs to be distinguished between *systems* and *displays*. As stated in the definitions above, McCrickard, Chewar et al. (2003), Stasko et al. (2004), Matthews, Rattenbury and Carter (2007) and Matthews, Dey et al. (2004) emphasise the unilateral nature of, according to their definition, *notification systems* or *peripheral displays*. As such, users do not provide input, but only receive information. The term system does not express this fact. Rather it generally refers to a 'set of things working together as parts of a mechanism (Oxford Dictionaries 2018d)'. The term display, on the other hand, accentuates the unidirectional attribute (Oxford Dictionaries 2018a).

Concluding, *peripheral awareness displays* are broadly defined as interfaces that are in the periphery of users' attention and convey information regarding one or more tasks of a user. Thereby, peripheral awareness displays may briefly move to the focus of attention if the urgency of the conveyed information increases.

5.2 Method

A systematic literature survey was performed to identify and analyse existing evaluation frameworks and methods for peripheral awareness displays. Whereas the literature review presented in Chapter 2 aimed at (a) providing a scientific foundation and (b) identifying research gaps, the review described in this chapter is performed to address the previously identified gaps and research questions, specifically those formulated in Subsection 2.9.2:

1. What are relevant criteria, dimensions, and methods for the design and evaluation of peripheral awareness displays?
2. How can these criteria, dimensions, and methods be implemented into a comprehensive design process for peripheral awareness displays?

Criteria are to be understood as principles a design can be judged against. Dimensions categorise different system designs and provide the basis for determining differences in weighting among criteria. Methods are procedures that can be applied to assess designs under consideration of criteria and dimensions. The literature was systematically reviewed following the five steps proposed by Khan et al. (2003): formulation of review questions, identification of relevant work, quality assessment, summary of findings, and interpretation of findings. Each step of the conducted systematic literature review is subsequently described in detail.

5.2.1 Formulation of Review Questions

The review adheres to the free form research questions that were previously formulated. The following constraints were derived to delimit the review:

- *Population:* Populations interacting with inanimate objects
- *Intervention/tool:* Peripheral awareness displays

- *Outcome:* Criteria, dimensions, and/or methods that aid the design, development and/or evaluation
- *Study design:* Case studies demonstrating the applicability of criteria

Thereby, different terms for peripheral awareness displays, such as notification displays, peripheral displays, or ambient displays (see Section 5.1), are to be taken into account.

5.2.2 Identification of Relevant Work and Quality Assessment

To include as many publications as possible, all databases accessible to the Loughborough University Library Catalogue Plus were searched, including Elsevier ScienceDirect, SpringerLink, Wiley Online Library, Taylor & Francis journals, SAGE, ACM Digital Library, IEEE Publishing, and PubMed.[1] No restriction to the year of publication was applied.

The search was performed using keywords corresponding to the previously identified structured review questions, with a focus on interaction medium, i. e. peripheral awareness displays (including alternative terms), and outcomes:

```
(display* OR system*) AND
(ambient OR peripheral OR notification) AND
(design* OR develop* OR eval*)
```

The primary search for article keywords led to a total of 1,645,917 results. An initial assessment indicated that the vast majority of results were unrelated to the review questions. To increase relevance, the search was reduced to comprise only the titles of articles, leading to a total of 342 results. The search results were then successively narrowed down by excluding terms related to other research areas, culminating in the following search code:

```
(display* OR system*) AND
(ambient OR peripheral OR notification) AND
(design* OR develop* OR eval*)
NOT (nervous~ OR cooler* OR energy OR cell* OR
blood OR infections OR intelligence OR living)
```

This resulted in 161 titles that were subsequently evaluated to assess their concordance with the structured review questions. In a first step, only those publications were selected for further analysis that addressed the interaction of humans with inanimate objects. Twelve publications fulfilled this requirement (Lang, Keith and Kavie 1994; McCrickard, Chewar

[1] A complete list of searchable databases can be accessed via `https://internal.lboro.ac.uk/info/library/uniwide/primocentral/`.

et al. 2003; Matthews, Rattenbury and Carter 2007; Miyamae et al. 2008; Jung and Schwartz 2008; Hoffmann, Wittke and Hahner 2009; Angelucci, Di Paolo and Tarantino 2009; Ongenae et al. 2012; Vatavu 2013; Alkhafaj and Fallahkhair 2014; Fortmann et al. 2015; Tang and Lee 2016). Next, it was assessed if the publications outlined criteria, dimensions, and/or methods that were applied for the design, development, and/or evaluation of peripheral awareness displays. Thereby, the existence of case studies relating to each of the described criteria functioned as a quality assessment measure.

This threshold was passed by three publications (McCrickard, Chewar et al. 2003; Matthews, Rattenbury and Carter 2007; Angelucci, Di Paolo and Tarantino 2009). A cited reference search was performed to identify related work, leading to an additional six publications that fulfilled the review criteria (Matthews, Dey et al. 2004; Pousman and Stasko 2006; Mankoff, Dey et al. 2003; Shami, Lshed and Klein 2005; Stasko et al. 2004; Matthews, Hsieh and Mankoff 2009). Overall, nine publications were selected for further analysis.

5.2.3 Summary of Findings

The identified publications are summarised in Table 5.1, whereby the order corresponds to the number of citations (see Section A.3 for a detailed description). As indicated by the first column, each publication either outlined criteria, dimensions, methods, or combinations of which. These types delineate general commonalities among the publications that provide the basis for a systematic analysis (see subsections 5.2.3.1 to 5.2.3.3).

As such, the Interruption, Reaction and Comprehension (IRC) framework proposed by McCrickard, Chewar et al. (2003) outlines criteria for the evaluation of peripheral awareness displays. Similarly, Matthews, Hsieh and Mankoff (2009) and Angelucci, Di Paolo and Tarantino (2009) contribute evaluation criteria to this review.

In contrast, Pousman and Stasko (2006) present design dimensions with which peripheral awareness displays can be categorised. Specifically, the dimensions are design choices that developers have to make prior to the development of the display. For instance, Pousman and Stasko (2006) propose the dimension *aesthetic emphasis*, whereas Matthews, Rattenbury and Carter (2007) present the criterion *appeal*. In this case, the dimension indicates the importance of the respective criterion and allows a weighting relative to other criteria, for instance *comprehension*.

The remaining publications present combinations of criteria, dimensions, and/or methods. Mankoff, Dey et al. (2003) adapted Nielsen's heuristics (Molich and Nielsen 1990), canonical usability principles, to establish a method for the low-cost evaluation of peripheral awareness displays. This involves the critique of a display by a small number of evaluators on the basis

Table 5.1: Summary of identified publications that serve as foundation for the development of a comprehensive evaluation framework and process for peripheral awareness displays. Publications are ordered by relevance, as indicated by number of citations (derived from Scopus, SpringerLink and Google Scholar, accessed 10 February 2019)

Type	Aspects	Authors	Origin	Citations
Method, criteria	Sufficient information design, consistent and intuitive mapping, match between design of ambient display and environments, visibility of state, aesthetic and pleasing design, useful and relevant information, visibility of system status, user control and freedom, easy transition to more in-depth information, peripherality of display, error prevention, flexibility and efficiency of use	Mankoff, Dey et al. (2003)	Heuristic evaluation (Molich and Nielsen 1990)	290
Dimensions	Information capacity, notification level, representational fidelity, aesthetic emphasis	Pousman and Stasko (2006)	Literature review	211
Criteria	Interruption, reaction, comprehension	McCrickard, Chewar et al. (2003)	Critical parameters (Newman 1997)	114
Dimensions, criteria	Abstraction, notification levels, transition	Matthews, Dey et al. (2004)	Attention model (Wickens, Hollands et al. 2013)	109
Dimensions, criteria	Personalisation, flexibility, number of information sources, accuracy, appeal	Stasko et al. (2004)	Literature review	100
Dimensions, criteria	Scope of use, criticality, appeal, learnability, awareness, effects of breakdowns, distraction,	Matthews, Rattenbury and Carter (2007)	Interviews, literature review	47
Method, criteria	Noticeability, comprehension, relevance, division of attention, engagement	Shami, Lshed and Klein (2005)	Literature review	32
Criteria	Learnability, error visibility, usefulness, user satisfaction, awareness, distraction	Matthews, Hsieh and Mankoff (2009)	Literature review	9
Criteria	Severity-dependence	Angelucci, Di Paolo and Tarantino (2009)	Literature review	2

of the derived heuristics and aims at the identification of usability issues. Supplementing the heuristic evaluation, Shami, Lshed and Klein (2005) introduce the Context of Use Evaluation of Peripheral Displays (CUEPD) which aims at evaluating peripheral awareness displays within their intended usage context. In addition to an evaluation procedure, CUEPD includes an evaluation questionnaire that serves as a further source for criteria. Lastly, Matthews, Dey et al. (2004), Matthews, Rattenbury and Carter (2007), and Stasko et al. (2004) present both evaluation criteria and dimensions.

5.2.3.1 Criteria

In the context of this review, the term *criteria* comprises all standards by which a peripheral awareness display can be judged. The following criteria were extracted from the previously identified publications (see Table 5.2 for a comparison and Section A.3 for a detailed description of each publication): appeal, comprehension, error visibility, learnability, usefulness, interruption, noticeability, reaction, urgency levels, and satisfaction. In the following, the criteria are defined in reference to their respective sources.

Appeal The criterion appeal delineates the hedonic qualities of a display, specifically its attractiveness (or aesthetics) and enjoyability. Corresponding evaluation criteria are present in four of the identified publications. As such, the heuristic 'aesthetic and pleasing design' proposed by Mankoff, Dey et al. (2003) informs this criterion. Further, the CUEPD questionnaire category *engagement* contains related questions. Specifically, it includes items that require participant responses to the perceived attractiveness and enjoyability of a display. Additionally, the frameworks presented by Matthews, Rattenbury and Carter (2007) and Matthews, Hsieh and Mankoff (2009) encompass similar requirements.

Comprehension It must be assured that users can cognitively process and make use of the information conveyed by the peripheral awareness display with a single glance. The criterion is mainly informed by the IRC framework (McCrickard, Chewar et al. 2003), for which it is a constituent parameter (see Subsection A.3.1). Further, the heuristics *sufficient information design*, *consistent and intuitive mapping*, and *match between system and real world* (Mankoff, Dey et al. 2003) as well as the CUEPD questionnaire category *comprehension* emphasise the relevance of this criterion for the evaluation of peripheral awareness displays.

Error visibility In case the peripheral awareness display malfunctions users must be made aware of the failure to avoid reliance on incorrect information and, consequently, to prevent commission or omission errors. This principle is informed by the similarly named aspect

Table 5.2: Identified evaluation criteria for peripheral awareness displays in reference to each source

Publications	Appeal	Comprehension	Error visibility	Learnability	Usefulness	Interruption	Noticeability	Reaction	Urgency levels	Satisfaction
Mankoff, Dey et al. (2003)	●	●	●	●	●	●	●	○	○	○
McCrickard, Chewar et al. (2003)	○	●	○	○	○	●	○	●	○	○
Matthews, Dey et al. (2004)	○	○	○	○	○	●	○	○	●	○
Stasko et al. (2004)	●	●	○	○	○	●	○	○	○	○
Matthews, Rattenbury and Carter (2007)	●	●	●	●	○	●	●	○	○	○
Shami, Lshed and Klein (2005)	●	●	○	○	●	●	●	○	○	○
Matthews, Hsieh and Mankoff (2009)	○	○	●	●	○	●	●	○	○	●
Angelucci, Di Paolo and Tarantino (2009)	○	○	○	○	○	●	○	○	●	○

● derived from publication ○ not present in publication

proposed by Matthews, Hsieh and Mankoff (2009) and the criterion *effects of breakdowns* (Matthews, Rattenbury and Carter 2007). Further, the heuristics *visibility of system status and state* relate to this principle (Mankoff, Dey et al. 2003).

Learnability Besides being comprehensible, peripheral awareness displays must be usable without further instructions after reasonable amounts of effort and time. *Comprehension* and *learnability* are recursive but not similar. While the comprehension of system features is a prerequisite for learnability, it does not guarantee it. Imagine a system with a series of simple steps that – each on their own – are easy to understand, but in their entirety difficult to remember and thereby not easy to learn. The similarly named criteria proposed by Matthews, Rattenbury and Carter (2007) and Matthews, Hsieh and Mankoff (2009) inform this principle. Further, the heuristics *sufficient information design*, *consistent and intuitive mapping*, and *match between system and real world* (Mankoff, Dey et al. 2003) highlight the importance of learnability.

Usefulness The information conveyed by the display must correspond with user needs. Specifically, the information must support users in achieving their goals. The heuristic *useful and relevant information* (Mankoff, Dey et al. 2003) and the CUEPD questionnaire category *relevance* (Shami, Lshed and Klein 2005) support this criterion.

Interruption By definition, peripheral awareness displays are designed to operate in the periphery of attention. However, the display must prompt users to reallocate their attention if required. Thereby, peripheral awareness displays must remain unobtrusive, i. e. non-interruptive, unless prespecified thresholds are exceeded. This criterion is supported by all selected publications.

Noticeability Users must be able to perceive the information conveyed by the peripheral awareness display without focal attention. The CUEPD questionnaire contains questions that evaluate if users noticed the display and if they were aware of the display while focusing on the primary task (Shami, Lshed and Klein 2005). Additionally, the heuristic *peripherality* (Mankoff, Dey et al. 2003) and the similarly named principles proposed by Matthews, Rattenbury and Carter (2007) highlight the importance of this criterion.

Reaction Reaction refers to the observable response of users to system stimuli. As described in the IRC framework (McCrickard, Chewar et al. 2003), displays must prompt users to rapidly and accurately react if the supported task requires it.

Urgency levels As previously outlined, the display should only attract attention if the situation requires it. As such, the display must be adaptive to several levels varying in urgency, i. e. salience. The severity levels proposed by Angelucci, Di Paolo and Tarantino (2009) and the notification levels outlined by Matthews, Dey et al. (2004) highlight the importance of this criterion. Following the attention categories discussed in Subsection 2.4.3, the peripheral awareness display can either be attended with divided, focal, or no attention. Matthews, Dey et al. (2004) mapped these categories to four notification (or urgency) levels: *Change blind* delineates insignificant information that should not attract attention, *make aware* describes somewhat important information that should result in divided attention between the primary task and the task supported by the peripheral awareness display, the levels *interrupt* and *demand action* both indicate increased urgency and require focal attention, whereby the latter level demands a user interaction.

Satisfaction Overall, the display should meet the users' expectations and result in a positive attitude towards the system. Matthews, Hsieh and Mankoff (2009) emphasise on the importance of this criterion for the general assessment of a peripheral awareness display.

The specification and significance of the presented criteria varies with the application. For instance, displays with a high emphasis on aesthetics and negligible usability requirements should be evaluated differently than high criticality displays for which aesthetics are of secondary importance. Thus, parameters are needed that allow the structured determination of optimum values and weightings for each criterion. Further, methods for the systematic evaluation of the criteria are required. Subsections 5.2.3.2 and 5.2.3.3 summarise the corresponding findings.

5.2.3.2 Dimensions

Dimensions afford the categorisation of peripheral awareness displays and enable the definition of optimum values and weightings for each of the previously derived criteria. In the following, the dimensions informed by the identified publications are defined (see Table 5.2). In reference to Pousman and Stasko (2006), peripheral awareness displays can be sorted into one of five bands for each dimension, ranging from low to high (low, somewhat low, medium, somewhat high, high).

Aesthetic emphasis This dimension describes the relative weight of criteria relating to the appeal of a peripheral awareness display. On one extreme, aesthetics may be crucial for the acceptance of a display in the consumer space. In contrast, aesthetics are presumably not as significant in professional environments such as aviation cockpits. Pousman and Stasko (2006) and Stasko et al. (2004) inform the selection of this dimension.

Information criticality Information criticality delineates the relative seriousness of the conveyed information. On one end of the scale, peripheral awareness displays can convey information that is relevant for survival. For instance, a display communicating the altitude of an aircraft. On the other end of the scale, the conveyed information may only serve the enjoyment of users. This dimension is related to the notification levels proposed by Matthews, Dey et al. (2004) and to the corresponding criterion *urgency levels*. However, high urgency levels do not necessarily correlate with a high information criticality.

Peripheral awareness displays may communicate content in a way that interrupts users and demands their attention (high urgency levels), but there may not be devastating consequences if users do not react (low criticality). Also, displays communicating content of high criticality will not always do this with the intention to grab the users' attention – unless the exceedance of specified parameters requires this. For instance, a peripheral awareness display communicating the altitude and speed of an aircraft may be unobtrusive as long as all values are within acceptable boundaries. However, if values exceed specified limits, the display should communicate this with a higher urgency, thereby interrupting users and potentially demanding their attention. This highlights that the urgency with which the content is communicated and the criticality of the information are somewhat independent from each other. While the criticality is a (mostly) static system parameter, urgency varies. Therefore, the presented model includes criticality as an overarching design dimension and considers different urgency levels as a dynamic parameter. For instance, with higher urgency levels it is more important that displays fulfil the criterion of interrupting the user.

Information capacity Information capacity refers to the distinct pieces of information that are to be conveyed. Peripheral awareness displays can, on the lower end, represent a single piece of information, for instance the velocity of the car, or several discrete information pieces, such as the position of nearby vehicles and their respective velocity.

Representational fidelity Representational fidelity refers to the degree of abstraction (Matthews, Dey et al. 2004), whereby increasing abstraction leads to faster perception, i. e. by making use of preattentive properties (see Subsection 2.4.3). However, more abstract designs impede intuitive understanding compared with more iconic representations (MacEachren, Roth et al. 2012; Pousman and Stasko 2006). The specification of this dimension aims at balancing this trade-off.

Personalisation This dimension concerns the degree of individualisation that the display allows. Thus, the display can either convey content from a predefined information source or lets users decide what information the display should communicate. For instance, a light display in a vehicle may be limited to showing the speed of the vehicle relative to the speed limit or it may be configurable to indicate other information like the distance to nearby road users.

Table 5.3: Identified dimensions for categorising peripheral awareness displays in reference to each source

Publications	Aesthetic emphasis	Information criticality	Information capacity	Representational fidelity	Personalisation
Pousman and Stasko (2006)	•	•	•	•	○
Matthews, Dey et al. (2004)	○	•	○	•	○
Stasko et al. (2004)	•	○	•	○	•
Matthews, Rattenbury and Carter (2007)	○	•	•	○	○

• derived from publication ○ not present in publication

5.2.3.3 Methods

Methods provide means by which peripheral awareness displays can be evaluated using the previously identified criteria (see Subsection 5.2.3.1). Two methods specifically targeted at peripheral awareness displays were extracted from the selected publications: the heuristic evaluation proposed by Mankoff, Dey et al. (2003) and CUEPD (Shami, Lshed and Klein 2005).

Heuristic evaluation Mankoff, Dey et al. (2003) introduce a method for the heuristic evaluation of ambient displays in terms of their effectiveness and usability. Their proposed set of heuristics adapts the set of ten usability heuristics which were originally proposed by Molich and Nielsen (1990) and subsequently extended (Nielsen and Molich 1990; Nielsen 1994a,a). Oxford Dictionaries (2018b) defines heuristics in the context of computing as a 'proceeding to a solution by trial and error or by rules that are only loosely defined'. An evaluation using heuristics involves several, potentially novice evaluators that review an interface based on images and a worded description (Nielsen and Molich 1990). Thereby, evaluators assess the compliance of the interface with established usability guidelines. This procedure was shown to discover between 40 and 60 % of known compliance problems of an interface (Mankoff, Dey et al. 2003). As indicated by a survey of usability practitioners, heuristic evaluation is regarded as one of the best methods for assessing interfaces, mainly due to being low cost, fast, and easy to apply (Vredenburg et al. 2002). The number of evaluators has to be selected under consideration of several factors, including the anticipated problem frequency, their impact, and visibility (Woolrych and Cockton 2001). For the evaluation of tasks with low complexity and diversity, five participants have been demonstrated to be sufficient for finding 80 % of the problems discoverable using heuristics (Nielsen and Landauer 1993). The heuristic evaluation procedure requires participants to detect issues with peripheral awareness displays relating to a predefined list of heuristics and assign a severity score to them, e. g. ranging from 1 (*low*) to 5 (*high*). The heuristics as selected and defined by Mankoff, Dey et al. (2003), ordered by issue coverage and including definitions from Nielsen (1994b) are described in Subsection A.3.5.

Context of use evaluation Shami, Lshed and Klein (2005) present the CUEPD as a complementary evaluation technique to the heuristic approach of Mankoff, Dey et al. (2003). CUEPD actively involves participants and incorporates the context of use through the building, enactment, and reflection of scenarios. As such, CUEPD requires the assessment of functional prototypes within the intended setting to take contextual influences on attention and perception into account. Figure 5.3 depicts the four steps required for CUEPD. In the first stage, scenario building, the designer interviews users to collaboratively construct

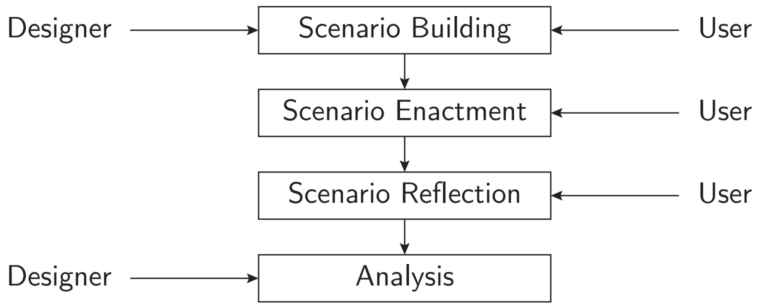


Figure 5.3: Conceptual model of CUEPD (Shami, Lshed and Klein 2005)

scenarios that represent the context of use. Thereby, the designer may ask users how they currently access the information that is to be communicated using the peripheral awareness display. The subsequent stage, scenario enactment, requires the user to perform a realistic primary task in the developed scenario while simultaneously accessing the peripheral awareness display. In the third stage, scenario reflections, participants are asked to fill in a survey to assess the peripheral awareness display against a set of predefined attributes. The final stage, analysis, numerically examines the results of the questionnaire.

Synergies The formative, heuristic evaluation of peripheral displays is an effective method for identifying usability problems early on in the development process. The technique merely requires the recruitment of 5 to 10 novice evaluators that check the proposed display concept against a set of heuristics. Thereby, annotated images of the interaction concept are sufficient, resulting in a rapid, low-cost evaluation technique. While useful for the detection of usability problems, the heuristic evaluation is not sufficient to evaluate peripheral awareness displays throughout the development process. In particular, attentional aspects such as interruption are difficult to review using low-fidelity prototypes (Matthews, Hsieh and Mankoff 2009). Contrasting the heuristic evaluation, the summative CUEPD relies on functional prototypes and is therefore more appropriate for later stages of the development process.

5.2.4 Interpretation of Findings

Within this subsection, the findings are interpreted in reference to complementing frameworks and criteria that are not specific to peripheral awareness displays, but applicable to the design of interactive products or systems in general. This aims at creating a comprehensive and structured framework for the evaluation of peripheral awareness displays. Further, the impact of each dimension on the weighting of criteria is discussed.

5.2.4.1 Complementing Criteria

While the systematic literature review has outlined criteria that are specific to peripheral awareness displays, it may have missed principles that are of relevance for displays in general. In Section 2.3, the concepts of usability and user experience were presented as general outcomes of product usage. By contrasting the identified criteria with usability and user experience models, it can be ensured that basic requirements are being considered.

Particularly the usability criteria (see Subsection 2.3.1.1) show significant overlap with those identified for peripheral awareness displays (see Table 5.2). For instance, learnability and usefulness map directly to the usability criteria of the same names.

The usability criteria safety, trustfulness, accessibility, and universality were not found in existing literature regarding peripheral awareness displays. Particularly accessibility and universality are relevant criteria to ensure that the intended user group is able to perceive the displayed information. Trustfulness mainly applies to (online) shops, whereas safety focuses on the consequences of wrongful user inputs (Seffah et al. 2006). As peripheral awareness displays are, in the context of this review (see Section 5.1), unilateral and are not designed to allow input, both requirements are not relevant for peripheral awareness displays. Further, the previously described criteria catalogue (see Subsection 5.2.3.1) lacks the consideration of the basic usability attributes effectiveness and efficiency. The comprehension, reaction, etc. does not in itself guarantee the appropriate use of the information. Thus, both attributes should be considered when evaluating peripheral awareness displays.

Besides criteria pertaining to the concept of usability, principles relating to user experience should be taken into account to achieve a holistic perspective. As such, the CUE model (see Subsection 2.3.2 and Figure 2.4) further extends the criteria catalogue by adding emotional responses and hedonic aspects. Moreover, the model can function as a framework for categorising the criteria into instrumental qualities, non-instrumental qualities, and overall system aspects.

The criteria that affect attention, namely interruption, noticeability, and reaction, are to be sorted into the instrumental features as they are characteristic to the support a system provides (Thüring and Mahlke 2007, p. 263). The degree to which displays, for instance, interrupt users and invoke specific reactions is independent of their aesthetic appeal. Thereby, the urgency level impacts the optimum values for attentional criteria (see Subsection 5.2.3.2).

As defined by Minge, Thüring and Wagner (2016), usability and the previously described subcriteria are also to be classed as instrumental qualities. In contrast to Seffah et al. (2006), the CUE model considers usability and usefulness as separate factors. As previously discussed (see Subsection 2.3.1.1), the criterion usability emphasises the effective and efficient use of a system. Information that is not useful cannot contribute to the appropriate use of a system in the sense of usability. Thus, the usefulness of a peripheral awareness display can be considered a subcriterion of usability.

In contrast to the categorisation proposed by Seffah et al. (2006) (see Subsection 2.3.1.1), the criterion satisfaction cannot be classified as instrumental since it is likely to be affected by non-instrumental qualities, such as aesthetics (Dillon 2001). Instead, satisfaction can be assessed as part of the component *appraisal of the system* (see Figure 2.4). Thus, the criterion usability includes the following subcriteria: effectiveness, efficiency, productivity, learnability, comprehensibility, safety, trustfulness, accessibility, and universality.

The visibility of errors completes the list of criteria regarding instrumental display qualities. Error visibility can be considered a distinct criterion because it is not significant as long as the system functions as intended and users may not even be confronted with errors at all. Due to their unobtrusive nature, peripheral awareness displays require errors to be highly visible. Otherwise, users may assume that the system is communicating non-urgent information and may therefore not notice when critical parameters are exceeded due to a reliance on the proper function of the display.

The remaining factors of the CUE model (Thüring and Mahlke 2007) and the meCUE questionnaire (Minge, Thüring and Wagner 2016) (see Subsection 3.2.3.6) that were not identified in the previous literature review are adopted in the criteria catalogue without further modification.

5.2.4.2 Optimum Values for Criteria and Relative Weighting

It is not desirable to aim for a maximisation of the outcome variables for all criteria. Peripheral awareness displays are, for instance, not necessarily to be rated higher with increasing interruption. Only if the situation requires it should the display interrupt the user. Thus, the attentional criteria – interruption, noticeability, and reaction – depend on the urgency level. Specifically, the degree to which users feel interrupted by the display in the execution of their primary task should be higher with increased urgency. Table 5.4 shows the optimum values of criteria depending on the urgency level. As discussed in Subsection 5.2.3.1, the notification level *change blind* should not interrupt the user and does therefore not require a reaction. Displays at the same level should, however, be somewhat noticeable. The next level, *make aware*, requires the display to somewhat interrupt users in what they are doing and may therefore invoke a reaction, corresponding to divided attention (see Subsection 2.4.3). At this point, the display must be noticeable. Finally, the two highest levels require maximum values for interruption, noticeability, and reaction, i. e. full attention.

Table 5.4: Optimum values of criteria relative to notification levels

Notification level	**Evaluation criteria**		
	Interruption	Noticeability	Reaction
Change blind	○	◐	○
Make aware	◐	●	◐
Interrupt	●	●	●
Demand action	●	●	●

Optimum values: ○ low; ◐ medium; ● high

Further, the importance of criteria differs depending on the ratings of two design dimensions. While information capacity, representational fidelity, and personalisation do not affect the relative importance of criteria, the criticality and aesthetic emphasis influence the importance of instrumental and non-instrumental qualities, respectively. As such, the attentional and overall instrumental criteria increase in importance with higher criticality. Similarly, the appeal-related, i. e. non-instrumental, qualities of the display become increasingly important with a higher emphasis on aesthetics.

It remains to be addressed, how the modules should be weighted. Hassenzahl (2005) argues that the weighting depends on the context in which the system is used. The specification of each design dimension results from the usage context and thus can be used as a basis for the weighting of criteria. As such, for systems that are highly critical and have little to no emphasis on aesthetics, instrumental qualities are likely to be more important than non-instrumental qualities. Similarly, for displays that have a high emphasis on aesthetics and communicate information of low criticality, non-instrumental qualities are more significant. As all components (instrumental, non-instrumental, emotional, overall appraisal) are relevant for determining the overall user experience (Minge, Thüring and Wagner 2016), they are assumed to be initially equally weighted. Against this background, Figure 5.4 illustrates the proposed weighting model for all components of user experience.

For instance, a similar specification of the dimensions information criticality and aesthetic emphasis leads to an equal weighting for all components. In contrast, for a display characterised by high criticality (5/5) and somewhat low aesthetic emphasis (2/5), instrumental qualities are of higher significance. As previously stated, five bands exist for each dimension (see Subsection 5.2.3.2). Thus, the weighting ratio can be defined as band A relative to band B. This leads to a ratio of 5:2 for the described example – and a weighting of 35.7 % for instrumental qualities and 14.3 % for non-instrumental qualities. The weightings of non-instrumental qualities w_{NI} and instrumental qualities w_{I} can, hence, be calculated as

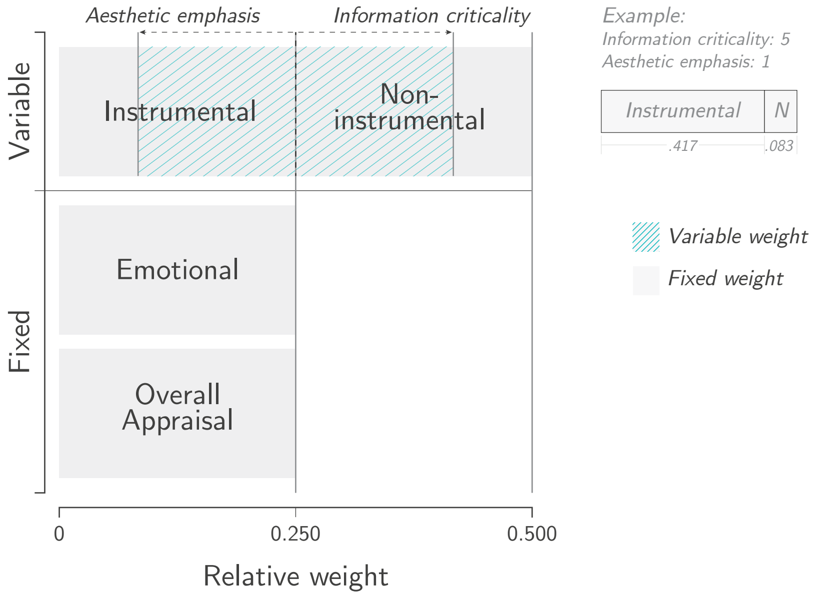


Figure 5.4: Proposed relative weights of evaluation components, the weight of instrumental and non-instrumental qualities is variable, depending on information criticality (1-5) and aesthetic emphasis (1-5). For instance, a high information criticality (5) combined with a low aesthetic emphasis (1) leads to a weighting ratio of 5:1, resulting in a weighting of 0.417 for instrumental qualities and 0.083 for non-instrumental qualities

shown in Equations 5.1 and 5.2. This results in a minimum relative weighting of 0.083 and a maximum relative weighting of 0.417.

$$w_{\mathrm{NI}} = \frac{0.5}{\frac{\text{Information criticality}}{\text{Aesthetic emphasis}} + 1} \tag{5.1}$$

$$w_{\mathrm{I}} = 0.5 - w_{\mathrm{NI}} \tag{5.2}$$

The aggregated UX score X can then be calculated from the individual scores multiplied with their respective weighting (see Figure 5.4):

$$X_{\Sigma} = \frac{X_{\text{usability,errors}} + X_{\text{attention}}}{2} \times w_{\mathrm{I}} + X_{\mathrm{NI}} \times w_{\mathrm{NI}} + (X_{\text{emotion}} + X_{\text{overall}}) \times 0.25 \tag{5.3}$$

The UX scores are relevant for the comparison of two or more peripheral awareness displays as they afford a systematic determination of the most suitable display within the respective usage context.

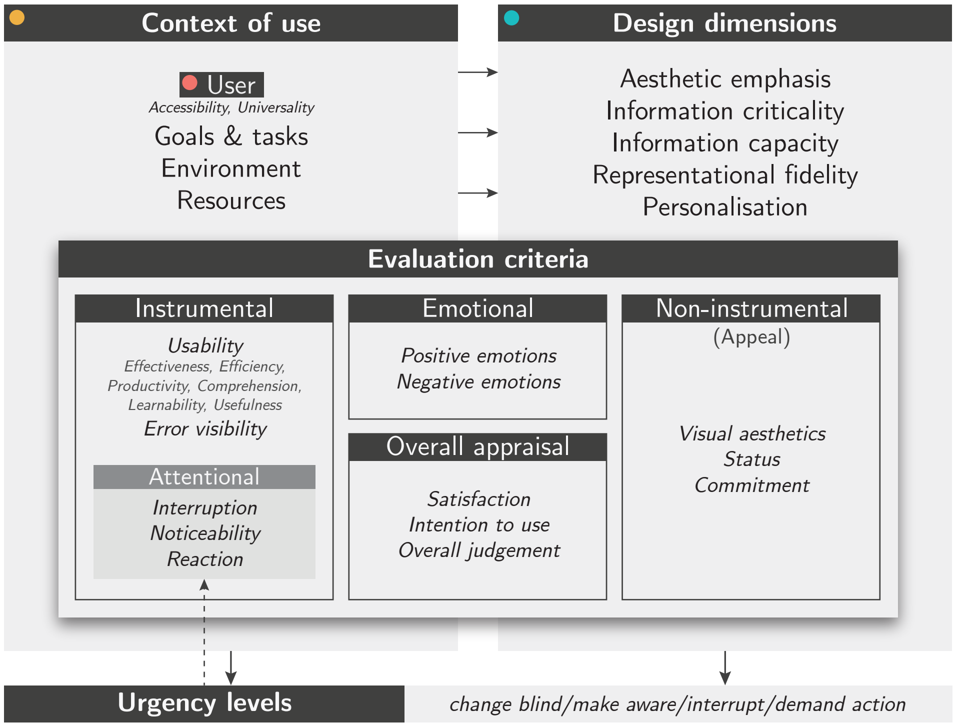


Figure 5.5: The PADE framework

5.3 Peripheral Awareness Display Evaluation Framework

Based on the derived evaluation criteria and dimensions, a comprehensive framework for the evaluation of peripheral awareness displays can be developed. For the evaluation of any interactive system, the consideration of the usage context is paramount (see Subsections 5.2.3.3 and 2.3). Consequently, the framework has to base any evaluation on factors that delineate the context of use, namely the users that interact with the display, their goals and tasks, the environment that a display is used in, and the available resources (see Figure 2.3).

Further, the developed display is to be classified along five distinct design dimensions to be able to determine optimum values for the evaluation criteria (see Subsection 5.2.3.2). Lastly, the criteria discussed in Subsection 5.2.4.1 are to be included.

Figure 5.5 depicts the proposed Peripheral Awareness Display Evaluation (PADE) framework. In contrast to existing frameworks, it takes a more comprehensive approach to the evaluation of peripheral awareness displays. As such, it incorporates the context in which the system is used while also considering how overarching design dimensions as well as urgency levels affect the optimum values for evaluation criteria. Existing models such as CUEPD include the context of use in the evaluation, but fail to combine this with an assessment of the impact of design dimensions. Further, prior models have a considerable focus on usability-related criteria without sufficiently considering non-instrumental features and emotional responses. The presented PADE framework integrates user experience criteria based on the CUE model and consequently balances the importance of instrumental and non-instrumental features as well as emotional responses.

Similar to CUEPD and CUE, the framework is implemented in the evaluation process as a questionnaire. The individual questionnaire items can be extracted from the previously analysed sources (see Table 5.5), whereby each subcriterion within a section is equally weighted. The sections are weighted according to Figure 5.4, whereby the attentional scores are to be allocated an equal weighting as the combined remaining usability and error visibility scores (learnability, comprehensibility, usefulness, error visibility). This is to account for the significance of attention-related aspects in peripheral awareness displays.

The questionnaire items belonging to the scale *negative emotions* are reversely worded when compared with the scale *positive emotions* (see Table 5.5). This needs to be accounted for through the following correction in coding: $Score_{reversed} = MIN + Range - Score$. For a 7-point Likert scale ranging from 1 to 7 this becomes: $Score_{reversed} = 8 - Score$.

The weighting and aggregation of all scores allows for the reduction of the evaluation to a single dimension ranging from 1 to 7 and helps to determine how well the display performs overall, which is particularly helpful when comparing two alternatives.

Thereby, it must be ensured that the questionnaire measures what it is intended to measure (validity) and that it generates reproducible results (reliability) (MacKison, Wrieden and Anderson 2010). Validity can be assumed as the questions are immediately derived from existing research. Reliability can be measured in terms of internal consistency (Trobia 2008). That is, the responses to items belonging to a single scale, e. g. learnability, must have a statistically significant degree of correlation with each other. Cronbach's alpha is commonly used to evaluate internal consistency (Trobia 2008). The value of alpha ranges from 0 to 1, whereby larger values indicate a higher internal consistency and thereby reliability. Alpha values of 0.70 and above are considered acceptable. In contrast, values that substantially exceed 0.90 indicate that one or more items are redundant and could be excluded (Trobia 2008). The evaluation of the questionnaire is performed as part of the user-based evaluation outlined in Subsection 6.2.3.

In addition to the questionnaire, the usability components effectiveness, efficiency, and productivity can be measured objectively, whereby the latter is only to be taken into account if there is a productive outcome. Further, reaction and emotions can be measured by recording the corresponding physiological and behavioural responses. Thereby it needs to be considered that peripheral awareness displays are typically not stand-alone products but rather a component of a larger system and are not at the centre of users' attention. This is particularly significant for the objective evaluation of emotions which are likely influenced by the primary task rather than the secondary task supported by the peripheral awareness display. Further, emotions may be the result of the nature of information and not of the display design (Scherer 2005). For instance, being notified that a takeover is imminent is highly stressful and will have an impact on emotional responses. Consequently, the evaluation of emotions based on physiological measurements cannot be recommended. The specific (quantitative) criteria for effectiveness, efficiency, productivity, and reaction depend on the usage context. As each criterion is likely on a different scale with unknown centre and end points, an implementation in the scoring system of the questionnaire is not practical. Consequently, measurements pertaining to these criteria should be reported alongside the questionnaire responses and UX scores. This constitutes a limitation of the presented framework implementation as the aggregated UX score (see Equation 5.3) does not encompass all criteria shown in Figure 5.5. However, the scores of related criteria in the questionnaire (see Table 5.5) imply the outcomes of effectiveness, efficiency, productivity, and reaction. On the one hand, high levels of learnability, comprehensibility, and usefulness promote effectiveness, efficiency, and productivity. On the other hand, appropriate levels of noticeability and interruption stimulate adequate reactions.

Table 5.5: PADE framework questionnaire

#	Sub-criterion	Questionnaire item	Source
		Instrumental qualities	
		Usability	
1	Learnability	It is quickly apparent how to use the display.	meCUE
2	Learnability	I was quickly able to effortlessly use the display.	Matthews, Rattenbury and Carter 2007
3	Learnability	It was easy to achieve the effortless use of the display.	Matthews, Rattenbury and Carter 2007
4	Comprehensibility	The display is easy to use.	meCUE
5	Comprehensibility	The operating procedures of the display are easy to understand.	meCUE

Table 5.5: (Continuation) PADE framework questionnaire

#	Sub-criterion	Questionnaire item	Source
6	Comprehensibility	I was able to understand the information just by glancing at it.	CUEPD
7	Usefulness	The functions of the display are exactly right for my goals.	meCUE
8	Usefulness	I consider the display extremely useful.	meCUE
9	Usefulness	With the help of this display I will achieve my goals.	meCUE
		Error visibility	
1	Errors	It is easy to notice failures of the display.	Nielsen 1994b
2	Errors	I was aware of the consequences of the failure.	Nielsen 1994b
3	Errors	I knew what to do when errors occurred.	Nielsen 1994b
		Attentional	
1	Overall	It was easy to perceive changes in urgency.	Matthews, Dey et al. 2004
2	Overall	I was able to shift my attention between my primary task and the display smoothly.	CUEPD
3	Overall	I was aware of the opportunity to access the display while performing my primary task.	CUEPD
4	Interruption (CB)	The display did not interrupt the primary task.	Matthews, Dey et al. 2004
5	Interruption (MA)	The display only slightly distracted me from the primary task.	Matthews, Dey et al. 2004
6	Interruption (I)	The display distracted me from the primary task.	Matthews, Dey et al. 2004

Table 5.5: (Continuation) PADE framework questionnaire

#	Sub-criterion	Questionnaire item	Source
7	Interruption (DA)	The display distracted me from the primary task so that I was not able to do anything else but monitor the road/vehicle.	Matthews, Dey et al. 2004
8	Noticeability (CB)	I could notice the display when I wanted to.	Matthews, Dey et al. 2004
9	Noticeability (MA)	I noticed the display.	Matthews, Dey et al. 2004
10	Noticeability (I,DA)	The display was clearly noticeable.	Matthews, Dey et al. 2004
		Non-instrumental qualities	
1	Visual aesthetics	The display is creatively designed.	meCUE
2	Visual aesthetics	The design looks attractive.	meCUE
3	Visual aesthetics	The display is stylish.	meCUE
4	Status	The display would enhance my standing among peers.	meCUE
5	Status	By using the display, I would be perceived differently.	meCUE
6	Status	I would not mind if my friends envied me for this display.	meCUE
7	Commitment	I could not live without this display.	meCUE
8	Commitment	The display is like a companion to me.	meCUE
9	Commitment	If someone took this display away, I would be devastated.	meCUE
		Emotions	
1	Positive emotions	The display exhilarates me.	meCUE
2	Positive emotions	The display relaxes me.	meCUE
3	Positive emotions	The display makes me feel happy.	meCUE

Table 5.5: (Continuation) PADE framework questionnaire

#	Sub-criterion	Questionnaire item	Source
4	Positive emotions	The display makes me feel euphoric.	meCUE
5	Positive emotions	The display calms me.	meCUE
6	Positive emotions	When using the display, I feel cheerful.	meCUE
7	Negative emotions	The display makes me feel tired.	meCUE
8	Negative emotions	The display annoys me.	meCUE
9	Negative emotions	When using the display, I feel exhausted.	meCUE
10	Negative emotions	The display frustrates me.	meCUE
11	Negative emotions	The display makes me feel passive.	meCUE
12	Negative emotions	The display angers me.	meCUE
		Overall appraisal	
1	Overall	Overall, I experienced the display as good.	meCUE
2	Satisfaction	I am satisfied with the display.	ISO 9241-210
3	Intention to use	If I could, I would use the display daily.	meCUE

5.4 Design Process for Peripheral Awareness Displays

The previous section addressed research question 1 (see Section 5.2) and established the PADE framework as foundation for the evaluation of peripheral awareness displays. It remains to be determined how this framework and the identified evaluation methods (see Subsection 5.2.3.3) can be implemented into a practical design process for peripheral awareness displays (see research question 2, Section 5.2). This section presents a development process that draws on reference processes and adapts them – under consideration of the previously identified framework and methods – to peripheral awareness displays.

5.4.1 Reference Processes

Similar to how usability and user experience criteria complemented the PADE framework, reference processes can guide the informed development of a procedure for the design of peripheral awareness displays. Two canonical reference processes were selected. To incorporate a HF perspective, the ISO-defined human-centred design approach for the development of interactive systems (ISO 2010) will be included. Contrasting this, the Double Diamond process provides a more creative approach to product development. Below, both processes will be introduced and analysed in terms of their implications for the development of peripheral awareness displays using the PADE framework.

5.4.1.1 Human-Centred Design for Interactive Systems

The standard ISO 9241-210:2010 (ISO 2010) describes a human-centred design approach for the development of interactive systems. Its comprised process 'aims to make systems usable and useful by focusing on the users, their needs and requirements, and by applying human factors/ergonomics, and usability knowledge and techniques'. Interactive systems are thereby defined as any 'combination of hardware, software and/or services that receives input from, and communicates output to, users (ISO 2010)'. Taken in isolation, peripheral awareness displays require no user input and are therefore not interactive. In the overall system context, however, peripheral awareness displays invoke user input to another device, for instance a steering wheel, and can therefore be considered part of an interactive system.

Figure 5.6 depicts the interdependence of human-centred design activities as described in the standard. The planning stage precedes the design activities and involves the allocation of resources and time, the selection of appropriate techniques, and the integration with the overarching project plan and other ongoing development processes. Four interconnected human-centred design activities form the outline of the design process:

1. Understanding and specification of the context of use: Descriptions of 'users, tasks, equipment (hardware, software and materials), and the physical and social environments in which a product is used (ISO 1998)'.
2. Specification of user requirements: User needs are to be derived under consideration of accessibility requirements (ISO 2006) and the intended usage context. The user requirements should be specified in terms that allow their consecutive testing and ensure internal consistency. Throughout the life of the project, the requirements are to be updated and verified by stakeholders.
3. Producing design solutions: Based on the established context of use, baseline evaluations and guidelines, potential solutions are to be designed. This involves the iterative design of interfaces under consideration of aspects pertaining to user experience.
4. Evaluation of design solutions: Potential design solutions are to be evaluated based on users' perspective throughout the design process. Particularly inspection-based evaluation and user-based testing are canonical user-centred evaluation approaches.

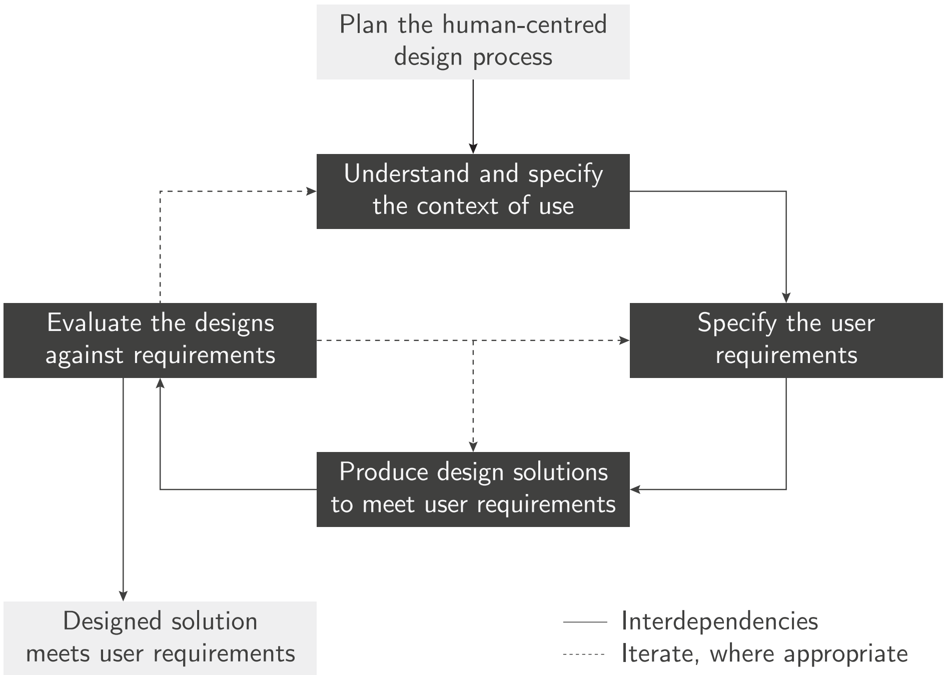


Figure 5.6: Interdependence of human-centred design activities (adapted from ISO 2010)

a) Inspection-based evaluation: Evaluation of the interface ideally performed by usability experts that is used to flag principal issues with interfaces before user testing.
b) User-based testing: Users are presented with a mock-up of the design concept for evaluation in the identified context of use, whereby the fidelity of the prototypes is to be increased throughout the process.

The approach to human-centred design closely resembles the CUEPD evaluation technique proposed by Shami, Lshed and Klein (2005). Both concepts emphasise the importance of considering the usage context during evaluation. Thereby, CUEPD serves as an example for a user-based testing method. The heuristics evaluation suggested by Mankoff, Dey et al. (2003), on the other hand, is an example of inspection-based evaluation.

The previously defined PADE framework can provide input for the specification of user requirements and thereby functions as a template for the second stage, specification of user requirements.

Overall, the human-centred design process outlined in the standard ISO 9241-210:2010 (ISO 2010) can be used as a model to merge and combine the presented frameworks and evaluation techniques for peripheral awareness displays. As such, several iteration cycles can make use of the advantages of the heuristic (formative) and CUEPD (summative) evaluation methods while integrating the identified user requirements.

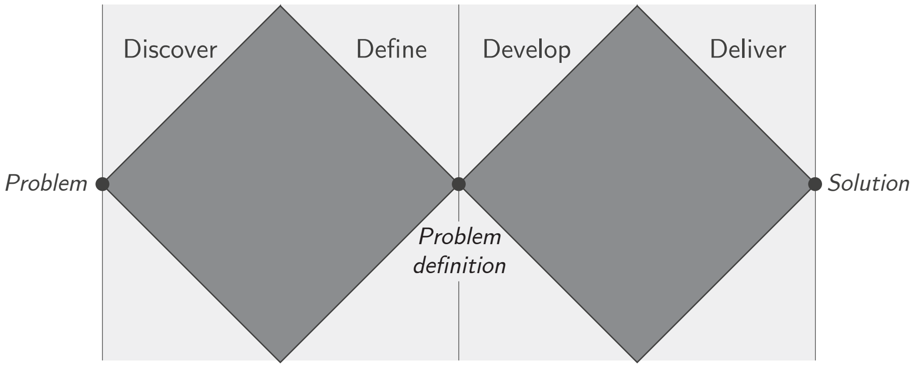


Figure 5.7: Schematised Double Diamond model (adapted from Design Council 2018)

5.4.1.2 Double Diamond Design Process

Created by the Design Council[2], the Double Diamond design process maps out the canonical creative approach designers across disciplines take to develop solutions (Design Council 2018). The process consists of four distinctive stages with two divergent and two convergent phases, creating the shape of a diamond (see Figure 5.7):

1. Discover: gathering insights and detailing the problem
2. Define: identifying the most critical problem under consideration of feasibility
3. Develop: iteration of concepts through prototyping and user testing
4. Deliver: finalise the resulting solution for product launch

For each phase, the Design Council describes a set of applicable methods. The *Discover*-stage focuses on research techniques to outline user characteristics and requirements, for instance observation, user diaries, surveys, and literature reviews. The second stage, *Define*, includes methods to evaluate and review the insights from the first stage to establish a concise design problem. This includes focus groups, the formation of assessment criteria, and customer journey mapping. The subsequent phase, *Develop*, centres around creative techniques to produce solutions and includes evaluation methods for the created solutions. This encompasses the definition of scenarios that capture the usage context and user profiles that resemble distinct user characteristics as well as physical prototypes. The final stage, *Deliver*, places an emphasis on evaluating high-fidelity prototypes or the final product against the identified requirements through user testing. Further, this includes all processes necessary to launch the product.

Compared with the previously presented process, the Double Diamond again emphasises on the importance of the usage context as well as the iteration of concepts and their

[2]The Design Council is a registered charity in the United Kingdom and a principal authority regarding the application of strategic design.

continuous evaluation. Overall, its distinct stages map to the design activities described in the standard ISO 9241-210:2010 (ISO 2010). As such, *Discover* and *Define* target similar outcomes as the design activities *understanding and specifying the context of use* as well as *specifying user requirements*, albeit with a larger emphasis on exploratory research. The two later stages, *Develop* and *Deliver*, exhibit similarities with the activities centred around the *production of design solutions* and their *evaluation*. Complementing the standard, the Double Diamond indicates during which stages divergent and convergent thinking are appropriate.

5.4.2 Comprehensive Process and Evaluation Procedure

Figure 5.8 summarises the resulting process for the development of peripheral awareness displays. It adopts elements from both the ISO and Double Diamond process and incorporates the previously described heuristic evaluation and user-based evaluation applying the PADE framework.

Similar to the Double Diamond process, diamond-shaped stages outline the divergent and convergent subphases. The first phase aims at gaining an understanding of user requirements and results in the specification of usage context and design dimensions as well as guidelines for the development of peripheral awareness displays. This includes a review of related work supplemented by primary research to address identified research gaps. Moving into the second phase, the guidelines serve as a starting base for initial ideation while the specified usage context and design dimensions regulate the development and evaluation.

Following Subsection 5.2.3.3, the evaluation procedure is to be divided into a heuristic evaluation for early stages and a user-based evaluation for later stages. For the heuristic evaluation, the process proposed by Mankoff, Dey et al. (2003) and the heuristics defined in Subsection A.3.5 can be used. The PADE framework provides the basis for the user-based

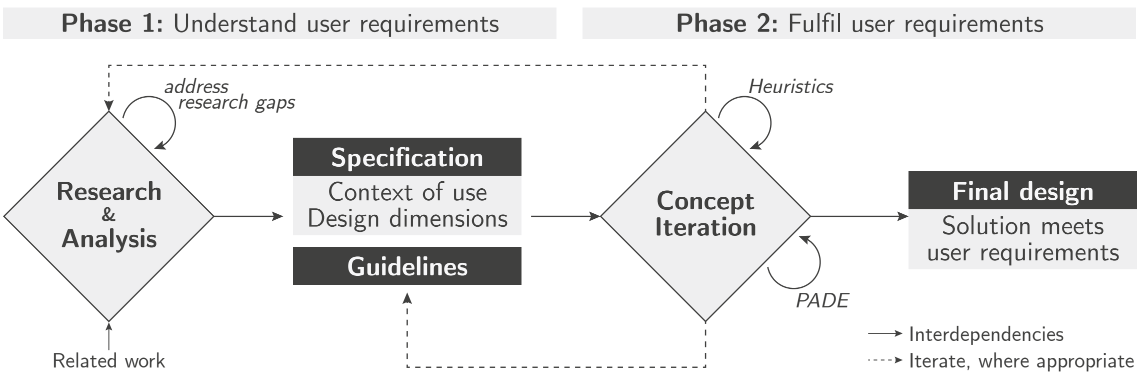


Figure 5.8: Development process for peripheral awareness displays based on ISO 9241-210:2010 (ISO 2010) and Double Diamond design process (Design Council 2018)

evaluation, whereby the previously specified design dimensions determine the weighting of instrumental and non-instrumental criteria.

The process follows the iterative approach described in ISO 9241-210:2010 (ISO 2010) (see Subsection 5.4.1.1). As such, limitations discovered during development or novel requirements identified as a consequence of evaluation can retroactively affect the design specifications. Section A.4 provides a step-by-step guide how the outlined process can be implemented to develop user-centred designs.

5.5 Summary and Conclusion

A novel end-to-end approach to the design and evaluation of peripheral awareness displays based on a consolidation of existing research has been developed. Addressing shortcomings of earlier work, the systematic literature review outlined in this chapter aimed at identifying criteria, dimensions, and methods that are relevant for the development and evaluation of peripheral awareness displays. The resulting PADE framework (see Figure 5.5) consolidates the findings into a comprehensive model. In contrast to previous work, the model provides a more holistic perspective on the development of peripheral awareness displays. As such, it includes a comprehensive set of evaluation criteria that include all components of user experience, i. e. instrumental and non-instrumental qualities, emotional reactions, and overall appraisal. Thereby, it encompasses all criteria that were identified in existing publications concerning the evaluation of peripheral awareness displays (see Table 5.2) and adds criteria external to this domain. Additionally, it provides an extensive questionnaire to operationalise the framework (see Table 5.5). Further, these criteria are interpreted against the backdrop of usage context and design dimensions, whereby the impact of the latter on the weighting of the criteria is quantified (see Figure 5.4). Thus, the model allows a numerical evaluation of peripheral awareness displays with respect to the task they support and the context they are used in.

Drawing on reference processes (see Subsection 5.4.1), the PADE framework is integrated into a development process for peripheral awareness displays that highlights its role during evaluation and, consequently, supports practical implementation.

Thus far, however, the developed framework and procedure is solely based on a review of the literature and lacks practical application. In the following chapter, the process is applied to the development of a peripheral awareness display for uncertainty communication.

CHAPTER 6

Application of Peripheral Uncertainty Communication

Abstract Peripheral awareness displays have been hypothesised to alleviate the attention-related derogative effects of visually conveying uncertainties (see Chapter 4). As a prerequisite for investigating this hypothesis, a display must first be developed that affords users to perceive the conveyed information without attentional focus. This chapter outlines the development of such a peripheral awareness display for uncertainty communication in the context of automated driving. As such, it continues where the previous chapter left off and applies the established PADE framework and development process (see Chapter 5). In a first step, design dimensions are specified and the usage context is analysed to identify suitable modalities for unobtrusive uncertainty communication. Under consideration of the multiple resource model (see Subsection 2.4.3), the peripheral visual and haptic channel are selected to prevent overburdening individual perceptual channels. Consequentially, visual and haptic variables are then evaluated in terms of their suitability for conveying states of and changes in uncertainties. In conjunction with the specified design dimensions, the results of this evaluation are subsequently used to develop an initial prototype display consisting of a light strip located on top of the centre console and a haptic seat. Through inspection-based, i. e. heuristics, and user-based testing applying the PADE framework the display is iteratively evolved until the final design meets the user requirements. In the subsequent chapter, the developed peripheral awareness display is then used to evaluate the implications of peripheral uncertainty communication. Aspects of this chapter, in part verbatim, are published in Kunze, Summerskill et al. (2018b,c, 2019b).

Background	Objectives	Method	Results	Implications
Using a digital instrument cluster to convey uncertainties requires operators to regularly change their gaze to perceive changes, leading to increased workload.	Development and evaluation of a peripheral awareness display for uncertainty communication using the previously defined development process and framework.	Systematic evaluation including inspection-based testing using heuristics and user-based testing applying the previously established PADE framework.	Peripheral awareness display for uncertainty communication that affords users to be aware of information while attending another task with focused attention.	The characteristics of the developed display render it suitable for overcoming the HF issues associated with visual uncertainty communication, which is evaluated hereafter.

Figure 6.1: Overview of Chapter 6

6.1 Research and Analysis

Following the process outlined in Figure 5.8, the first development phase aims at gaining an understanding of user requirements. In an initial step, the intended usage context and design dimensions are specified to set limits for the development. Based on the defined context of use and the specified dimension bands, related work is reviewed to identify guidelines that inform and delimit the ideation process (see Section 6.2). This is supplemented by two experiments that address shortcomings of existing work. The first phase concludes with a specification of overall guidelines that serve as a foundation for the second phase, the fulfilment of user requirements.

6.1.1 Context of Use

The context of use includes four aspects (see Figure 5.5): users, environment, goals & tasks, and resources. In the following, each element is defined in regard to the intended application.

6.1.1.1 Users

The peripheral awareness display is aimed at production vehicles, therefore all licenced drivers must be able to perceive the information it conveys. To be legally allowed to drive on public roads, the UK government requires drivers to (a) be in possession of a valid driving licence, (b) be at least a specified minimum age (typically 17), and (c) meet the eyesight requirements. The latter demands drivers (with corrected vision if necessary) to be able to read a licence plate from a distance of 20 m and have a visual acuity of 0.5 (Snellen scale) or better. There are no requirements in regard to colour vision.

6.1.1.2 Environment

The peripheral awareness display is to be located in the cockpit of a vehicle equipped with a driving automation system. As discussed in Subsection 2.5.6, SAE level 3 (SAE International 2018) is the most challenging to design for and users of vehicles classified with this level of automation likely benefit most from uncertainty communication. Further, the visual uncertainty display used in Chapter 4 was designed for SAE level 3 (SAE International 2018). Thus, choosing the same target level for the peripheral awareness display facilitates a comparison between the two display types, laying the foundation for evaluating if peripheral uncertainty communication reduces operator workload relative to visual displays (see Chapter 7).

6.1.1.3 Goals and Tasks

Safety-related goals are paramount for drivers and are prioritised over competing goals relating to, for instance, fuel or time saving (Dogan, Steg and Delhomme 2011). To fulfil safety goals, SAE level 3 requires drivers to be fallback-ready at any time (see Subsection 2.5.6). However, even without driving automation, crash statistics indicate that drivers regularly perform NDRTs while driving, leading to increased distraction that contradicts the stated importance of safety (Horberry et al. 2006). In 2017, 2,994 distracted drivers were involved in fatal crashes on US roads (NHTSA 2018). Distraction is most significant for teens aged 15 to 19: 9 % of the total 2,526 teen casualties were affected by distraction (NHTSA 2018). In a survey from the same year conducted by the AAA Foundation for Traffic Safety (2018) and involving 2,613 US licenced drivers, 60.5 % of participants reported to have talked on a hands-free mobile phone, while 49.1 % indicated to have used a hand-held mobile phone during the past month. At any given moment, an estimated 2.9 % of drivers use hand-held mobile phones while driving, despite it being illegal (National Center for Statistics and Analysis 2019). 44.9 % reported to have read an email or text message within the last month, whereas 34.6 % indicated to have sent a text message or email in the same time period (AAA Foundation for Traffic Safety 2018). As outlined in Subsection 2.5.2, the engagement in NDRTs is likely to increase with higher automation levels. The results from a recent survey involving 2,932 US licenced drivers indicates what kind of tasks users are willing to perform during activated driving automation (see Figure 6.2) (Erie Insurance 2017). In agreement with data from the previously described survey, talking on the phone (45 %), texting (34 %), and checking/sending emails (34 %) are prevalent NDRTs.

6.1.1.4 Resources

As discussed in Section 2.4, human information processing is affected by the limited availability of attentional resources. Driving manually – or monitoring an uncertainty display during activated driving automation – while performing an NDRT constitutes a multitasking situation. The multiple resource model (Wickens, Hollands et al. 2013) outlined in Subsection 2.4.2 highlights that concurrent tasks should address different resource pools to prevent interference and task performance decrements. Consequently, the peripheral awareness display should address perceptual channels that are not preoccupied with NDRTs. As indicated by the previously described survey results about activities during automated driving (see Figure 6.2), the (focal) visual channel (texting, checking and sending emails, reading, etc.) and the auditive channel (talking on the phone, watching videos, etc.) are

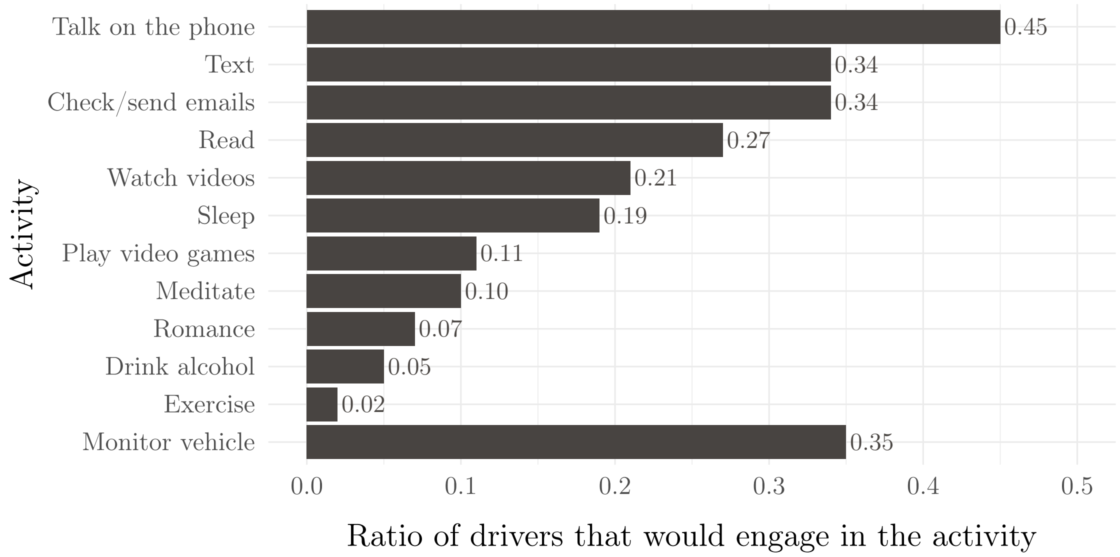


Figure 6.2: Results from a survey involving 2,932 US licenced drivers that were asked what tasks they would perform during activated driving automation (Erie Insurance 2017)

likely addressed by NDRTs and should therefore be avoided by the uncertainty display to prevent overburdening users.

The derogative effects of addressing similar perceptual channels in multitasking situations are confirmed by the outcomes of the experiment outlined in Chapter 4. The simultaneous performance of a visually demanding NDRT and the monitoring of the visual uncertainty display led to significant workload increments relative to the control group (see Table 4.3).

Modalities that are likely not occupied but familiar in the context of driving are the peripheral visual and the tactile channel. Both are characterised by fast perception dwell times and therefore support a rapid processing of the conveyed information (see Table 2.1). Relying on only one of the channels should, however, be avoided. The discussion in Subsection 2.4.3.1 indicated that humans are poor at noticing unexpected events within their field of vision, as highlighted by phenomena such as change blindness and inattentional blindness. Thus, when addressing solely the (peripheral) visual channel, even salient uncertainty changes may be missed. Further, users with colour blindness may not be able to perceive the information conveyed by lights. Relying exclusively on tactile stimuli is equally problematic. Specifically, vibro-tactile feedback leads to attentional spikes that immediately attract the attention of users (Kahol et al. 2006). Conveying uncertainties solely based on tactile feedback would therefore be unsuitable for conveying content in the periphery of attention. Instead, both channels should be addressed to make use of their respective benefits, resulting in a multimodal interface. As discussed in Subsection 2.5.3.3, multimodal cues are particularly useful for conveying an increase in urgency and were shown to improve reaction time in the context of TORs. Thus, peripheral visual cues can be employed to provide unobtrusive feedback for lower urgency levels, whereas a combination of both peripheral visual and tactile signals can indicate changes towards higher urgency levels. Further, addressing the tactile channel may overcome issues associated with attentional tunneling (see Subsection 2.4.2). As such, the provision of tactile feedback may interrupt the attentional fixation, particularly if the tactile stimuli are not constantly presented but rather only to indicate changes (Self et al. 2008). Thus, addressing both the tactile and the peripheral visual channel can prevent overloading users and attentional tunneling.

Table 6.1 summarises the context of use specifications for each of the four discussed elements. The specified elements are hereafter used to determine the bands of each design dimension and to guide the review of related work.

Table 6.1: Specification of the usage context for peripheral uncertainty communication in automated driving

Element	**Specification**
User	Drivers in possession of a valid driving licence
Environment	Vehicle equipped with SAE level 3 automated driving system
Goals and tasks	Performance of visual and auditive NDRTs while remaining fallback-ready
Resources	Peripheral visual and tactile channels likely unoccupied

6.1.2 Specification of Design Dimensions

As depicted in Figure 5.5, the following five dimensions characterise peripheral awareness displays: aesthetic emphasis, information criticality, information capacity, representational fidelity, and personalisation. Each of these design dimensions can be divided into five bands, ranging from low to high (see Subsection 5.2.4).

The importance of aesthetics in regard to the uncertainty display can be derived from the overall significance of aesthetics when buying a car. In a survey involving 346 German car buyers, 41 % of participants named *aesthetics* as a reason for buying a specific vehicle (see Figure A.5). As such, the aesthetics of a vehicle are the fourth most prominent selection criterion for buying a new car, after *value for money* (48 %), *comfort* (47 %), and *safety* (41 %). Further, brand affection is more influenced by emotional design aspects than the functional value of the vehicle (Kumar, Townsend and Vorhies 2015). Therefore, the dimension *aesthetic emphasis* can be specified as somewhat high, i. e. 4 out of 5.

With respect to the second dimension, information criticality, the uncertainty display can be specified as high (5/5) as it conveys information relevant for driving safety and, consequently, survival. As only the uncertainty information is to be conveyed, both information capacity and personalisation are to be specified with minimum values (1/5).

Finally, the representational fidelity of the conveyed information needs to be specified. On the lower end of the scale, the display could employ abstract signifiers such as colour hue. On the upper end, one could use indexical signifiers such as photographs (Pousman and Stasko 2006). For the application at hand, rapid information processing is paramount to lower cognitive workload and improve reaction times. With lower informational complexity, abstract signifiers are more rapidly processed (see Subsection 2.4.3) (MacEachren, Roth et al. 2012; Healey and Enns 2012) and are therefore more suitable for the application at hand. Thus, the representational fidelity is specified as low, i. e. 1 out of 5.

6.1.3 Related Work

The review of related work focuses on the identification of guidelines for the communication of information using the previously identified, likely unoccupied peripheral visual and tactile channels (see Table 6.1). Several research projects have provided guidelines for conveying information by addressing the specified channels. The following review of related work first examines the use of lights as a means for visually conveying information that is perceptible without focal attention. Thereafter, research investigating tactile feedback is examined.

6.1.3.1 Peripheral Light Displays

Faltalous et al. (2018) used a light strip attached to a tablet computer positioned on the centre console to convey the reliability of an automated driving system. The colour of the light strip changed gradually from green, indicating a high reliability, to red, suggesting a potentially imminent TOR. The results of post-experiment semi-structured interviews revealed that the majority of participants struggled to distinguish between different uncertainty degrees and would prefer a reduction to less levels, for instance limited to red, yellow, and green, thus forgoing gradual changes in colour. Further, participants suggested the use of multimodal feedback, whereby visual feedback was perceived as least annoying and auditory and haptic stimuli are to be reserved for higher urgency levels.

Matviienko et al. (2015) investigated different light patterns for conveying everyday information such as temperature or jogging speed and derived several guidelines for peripheral light displays. Supporting the findings of Faltalous et al. (2018), the authors discovered that system states are to be conveyed stepwise rather than using colour fades. Further, participant responses indicated that red blinking light was most suited for conveying urgent information.

Whereas the previously described publications used traffic light colours, Ou et al. (2012) investigated the impact of warmer (more reddish) and cooler, i. e. bluish, colour hues. The results indicate that warmer colours are perceived as more *active*, whereas cooler colour hues are more *passive*.

Grounded in the context of assisted living, Davis et al. (2017) provide additional guidelines for encoding activity information using lights. In line with Ou et al. (2012), red is considered appropriate for indicating activity, whereas blue relates to resting periods. Further, the light should not be positioned in the immediate FOV, but rather in the periphery.

Schaller, Schiehlen and Gradenegger (2008) proposed the use of an illuminated ring around the steering wheel centre (see Figure 6.3) to communicate the status of the lateral component of a traffic jam assistance system. This peripheral display lights up green when the system is activated. Van Den Beukel and Van Der Voort (2014) proposed to extend the application of this display by indicating the status of the steering assistant with (unspecified) different colours and by flashing lights for imminent takeovers. Similarly, the Volvo XC90 uses red flashing lights to communicate a collision warning and to prompt an immediate user reaction (Manca, Winter and Happee 2015).

Mueller et al. (2014) proposed a peripheral awareness display attached to the perimeter of a tablet to show off-screen points of interest. As part of their development process, they investigated how criticality can be encoded using lights. In line with other research (Matviienko et al. 2015; Faltalous et al. 2018; Ou et al. 2012), the results indicate that particularly the traffic light metaphor (green to red) and the temperature metaphor (blue/cold to red/warm) are learned representations for changes in criticality. Further, a pulsing red light was perceived as most urgent.

In a driving context, the majority of approaches that aim to convey information through changes in lighting use light strips, which consist of several aligned LEDs placed on a (flexible) band. Light strips have the advantage of being easily fitted into existing cockpits without affecting their overall look. For instance, Borojeni, Chuang et al. (2016) positioned

Figure 6.3: Peripheral light display indicating that the lateral component of a traffic jam assist is activated (Schaller, Schiehlen and Gradenegger 2008)

a light strip for conveying TORs behind the steering wheel. Loecken, Heuten and Boll (2015) located a light strip on top of the door panel to support lane change decisions.

It remains to be determined where such a light strip should be positioned in a vehicle cockpit. Löcken et al. (2013) explored nine different locations for light strips in automotive cockpits in terms of their perceptibility and preference by participants (see Figure 6.4). Locations in the instrument cluster (D) and on top of the centre console (F, H) performed best for both factors.

Further, ISO 9241-20:2008 (ISO 2006) specifies the accessibility requirements a system should fulfil. As previously discussed, users with limited colour vision can obtain a driving licence and are therefore potential users of the peripheral uncertainty display. The selected light coding should therefore be redundant in order to make the system accessible to users suffering from colour blindness (ISO 2006). Additionally, passengers with photosensitive epilepsy may react to flicker, thus the maximum flash rate must not exceed 3 Hz (Epilepsy Society 2018).

Based on the described existing body of research, the following guidelines for the design of a peripheral light display can be derived:

1. System states, such as uncertainty levels, should be conveyed using distinct (colour) steps rather than fades
2. High urgency is conveyed with red, highest urgency with pulsing red light
3. Traffic light metaphor and temperature metaphor are learned representations for conveying changes in urgency
4. Red is perceived as more active, whereas blue is perceived as rather passive
5. Peripheral light display should not be positioned in direct FOV
6. In an automotive context, peripheral light displays are typically implemented as light strips
7. Light strip should be positioned in the instrument cluster or on top of the centre console
8. Implement redundant coding to make light accessible by all potential users
9. Flash rates must not exceed 3 Hz

6.1.3.2 Vibro-Tactile Displays

In an automotive context, approaches using tactile feedback have focused on stimuli transmitted through the driver's seat (see Subsection 2.5.3.3). Using the seat for tactile feedback has several advantages over other vehicle components such as the steering wheel. Firstly, the driver is constantly in touch with the seat. In contrast, drivers likely do not have their hands on the steering wheel during activated driving automation but are rather

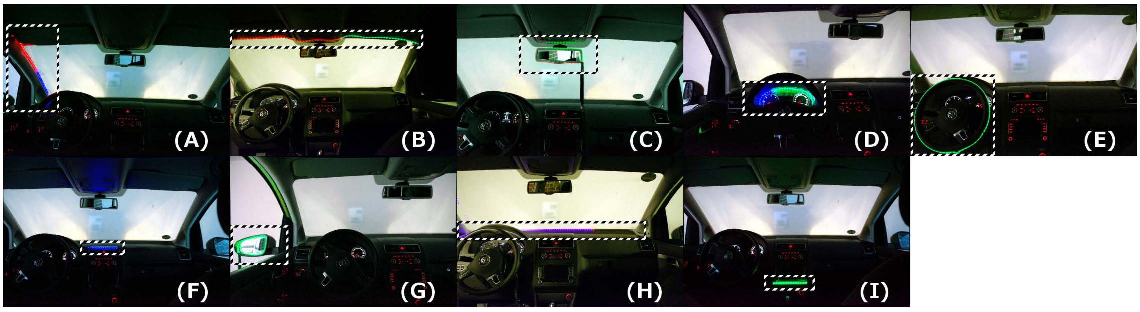


Figure 6.4: Different locations of light strips in a vehicle cockpit evaluated by Löcken et al. (2013)

engaged in NDRTs (see Subsection 2.5.2). Further, the seat is in touch with a large body area, allowing for more complex vibration patterns as opposed to, for instance, implementing the vibration in handheld devices.

In an automotive context, the majority of vibro-tactile displays use eccentric rotating mass vibration motors that consist of a direct current (DC) motor to which an offset mass is attached (ISO 2011b). The rotation of the asymmetric mass results in a centrifugal force that displaces the motor. With a sufficient rotation frequency, the displacement of the motor is perceived as vibration. The vibration can be intensified through increases in amplitude and frequency, whereby both are inseparably linked (Precision Microdrives Limited 2019).

It remains to be addressed how uncertainties can be conveyed using suchlike motors. Self et al. (2008) suggest a series of vibro-tactile characteristics for use in military environments: size (small or large), shape (circle or cross), orientation (vertical or horizontal), position, moving patterns, frequency, amplitude, rhythm, and waveform (square, triangular, sine waves, or saw tooth). Ji, Lee and Hwang (2011) derived the following characteristics for application in a vehicle seat and aimed to identify guidelines for each: frequency, amplitude, position, moving pattern, and rhythm. For this, they placed eccentric rotating

Table 6.2: Numerical recommendations for vibro-tactile characteristics in an automotive seat (Ji, Lee and Hwang 2011)

Characteristic	**Variable**	**Experimental results**
Amplitude	Appropriate intensity	Seat pan: 2.02 G to 2.65 G Back support: 2.65 G to 3.38 G
Moving pattern	Minimum distance between vibrations	Seat pan: 8 cm Back support: 9 cm
Rhythm	Satisfaction	Highest: 3 s vibration separated by 0.5 s interval Lowest: 3 s vibration separated by 2 s interval

mass vibration motors beneath a layer consisting of 1.2 mm leather and 8.8 mm sponge. In a series of four experiments, they identified the recommendations shown in Table 6.2 and Figure 6.5. It needs to be noted, however, that the recommendations regarding an appropriate vibration intensity cannot be transferred to other mechanical systems without further ado because they are based on a normalised amplitude. Thus, the indicated G values can only be used as rough guidelines.

Concerning accessibility, ISO 9241-920:2009 (ISO 2009) specifies that humans are only capable of distinguishing between frequencies that fall within the rang of 10 Hz to 600 Hz. Further, stimuli must not go below a duration of 10 ms to be perceivable.

In reference to the discussed work, the following guidelines for the design of a vibro-tactile display can be derived:

1. Use of eccentric rotating mass vibration motors positioned in the driver's seat
2. Vibro-tactile characteristics that can be varied using a vehicle seat are amplitude, position, moving pattern, and rhythm
3. Numeric recommendations as shown in Table 6.2
4. Position and direction of vibration motors depicted in Figure 6.5
5. Vibration frequency must be in a range from 10 Hz to 600 Hz to allow a perception of differences in frequency
6. Vibration stimuli should last at least 10 ms to be perceivable and should be separated by gaps of at least 10 ms

6.1.3.3 Research Gaps

The review of related work has provided a series of guidelines for both the use of a peripheral light and a vibro-tactile display. Nonetheless, it remains to be addressed how uncertainties should be conveyed using the specified channels. Specifically, it needs to be evaluated what light and vibro-tactile variables are best suited for communicating uncertainties. To address this research question, two laboratory experiments were conducted, one for each perceptual channel.

6.1.4 Evaluation of Light Variables

As a prerequisite for uncertainty communication using a peripheral light display, suitable light variables must allow users to confidently distinguish several levels of uncertainty. Without being able to perceive differences between several levels, users are not able to use the conveyed information. According to the specification of the dimension representational fidelity discussed in Subsection 6.1.2, the uncertainties are to be conveyed using abstract signifiers in order to afford the rapid processing of the conveyed information. Increasing

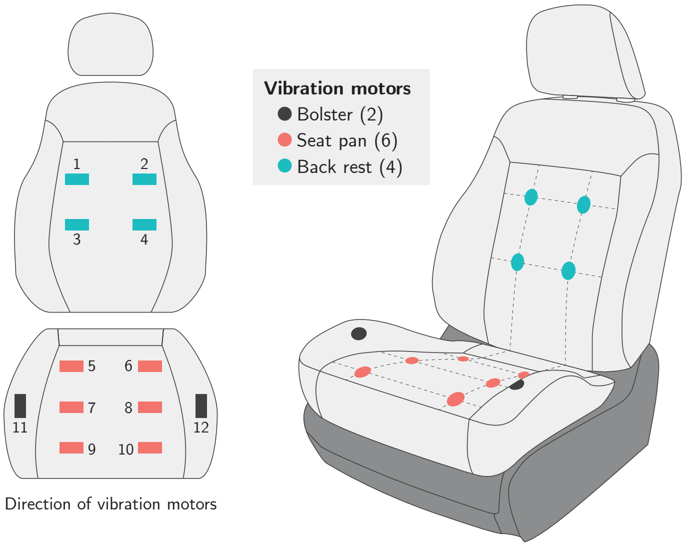


Figure 6.5: Position and direction of eccentric rotating mass vibration motors on vehicle seat (Ji, Lee and Hwang 2011; Chang, Hwang and Ji 2011; Kunze, Summerskill et al. 2018c)

abstraction, however, poses the problem that users might not be able to connect the signifier, i. e. the light variable, with the signified, i. e. the uncertainty level. Whereas literal descriptions or images may explicitly state the current level of uncertainty, abstract signifiers require users to link the stimuli with specific information, i. e. to make sense of what is being conveyed. Thus, it needs to be evaluated what abstract light variables are logical for conveying uncertainties.

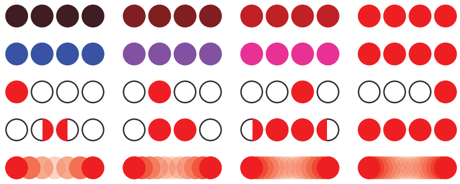
Further, peripheral awareness displays should increasingly interrupt users with higher urgency levels. Therefore, it is of interest how interrupting the light variables are perceived. Overall, this leads to the following research questions:

1. Which light variables afford users to confidently distinguish between several levels?
2. What light variables are most logical for conveying uncertainties in the context of automated driving?
3. What light variables are perceived as most interrupting?

6.1.4.1 Selection and Specification of Light Variables

It remains to be determined which light variables are to be selected for the evaluation. To achieve maximum abstraction, the light variables can be derived from the basic building blocks for visual representations, i. e. visual variables (see Subsection 2.8.2.2 and Figure 2.17). The following variables were selected for further investigation: brightness, colour hue, size, position, movement, and pulse.

Table 6.3: Schematic representation of the levels implemented for each light variable (Kunze, Summerskill et al. 2018b)



	Urgency levels			
	1	2	3	4
Brightness				
Colour hue				
Position				
Size				
Movement				
Pulse	0.75 Hz	1.50 Hz	2.25 Hz	3.00 Hz

As a basis for the evaluation, specific instances for each variable need to be determined (see Table 6.3). Distinct stages are thereby preferred over gradual changes (see Subsection 6.1.3.1). In reference to the urgency levels depicted in Figure 5.5, four instances were chosen for each variable. Following the guidelines described in Subsection 6.1.3.1, red (RGB: 255, 0, 0) was used as a base colour as it is best suited for conveying critical information.

Brightness was varied linearly from RGB(64, 0, 0) to RGB(255, 0, 0). Referencing the specified guidelines (see Subsection 6.1.3.1), two metaphors were eligible for selecting colour hues: traffic lights (green, amber, red) and temperature (red to blue). For reasons of accessibility, specifically due to the prevalence of protanomaly and deuteranomaly, traffic light colours were avoided. The uncertainty display indicates when an action, i. e. monitoring or takeover, is required by users. Thus, for low uncertainties, the display should convey that the user can remain passive while the display should communicate the need for an action in case of higher uncertainty. This active-passive indication is intuitively conveyed by red and blue (see Subsection 6.1.3.1). Consequently, four levels ranging from blue (RGB: 0, 0, 255) to red (RGB: 255, 0, 0) were selected. Intermediate values were selected by linearly adjusting the *hue* component of the hue, saturation, lightness (HSL) colour codes (see Table 6.4). ISO 9241-303:2011 (ISO 2011a) specifies that the differences between two colours must exceed a value of $\Delta E^*_{uv} = 20$ to be distinguishable. ΔE^*_{uv} is a numeric indicator for perceptive differences in colour proposed by the International Commission on

Table 6.4: Colour codes for hue variable; ΔE^*_{uv} and ΔE^*_{00} indicate the colour difference between the colour in the same row and the row below

	RGB	**HSL**	**CIE-L*uv**	ΔE^*_{uv}	ΔE^*_{00}
Level 1	0, 0, 255	240, 100, 50	32.297, -9.405, -130.342	42.325	22.202
Level 2	170, 0, 255	280, 100, 50	46.667, 30.404, -130.000	126.780	29.415
Level 3	255, 0, 170	320, 100, 50	56.256, 122.283, -43.170	96.637	34.219
Level 4	255, 0, 0	360, 100, 50	53.241, 175.015, 37.756	-	-

Illumination (CIE). As shown in Table 6.4, the differences between all levels exceed this value.

The variables size and position were varied linearly based on the number of activated LEDs (RGB: 255, 0, 0). In accordance with the derived guidelines, the flashing rate for the animation-based variables pulse and movement was limited to 3 Hz to prevent seizures as a consequence of photosensitive epilepsy. Starting from this upper frequency, both the pulse rate and movement speed were then, similarly to the other variables, linearly reduced to one fourth of the maximum value. Specifically, this results in the levels 0.75 Hz, 1.50 Hz, 2.25 Hz, and 3.00 Hz for the pulse variable. Movement was implemented as an animation starting from the centre and moving to the outer edges of the light strip (see Table 6.3). Thereby, the activation time for each LED was limited to 333.33 ms for the highest level.

6.1.4.2 Method

The formulated research questions (see p. 159) were investigated in a laboratory-based experiment. Ethical approval was granted by the Ethics Approvals Sub-Committee of Loughborough University.

Participants A total of 25 participants (7 female) with an average age of 32.640 years (SD=10.000, range=19-55) volunteered to participate in the experiment. All participants had normal or corrected-to-normal vision with no diagnosed visual impairments such as colour blindness.

Experimental Design The experiment employed a within-subjects design, whereby each participants experienced every variable in an overall balanced order. To assess research question 1, i. e. how confidently participants are able to distinguish between different levels of each variable, fuzzy membership functions were used. Originating in linguistics, fuzzy membership functions indicate to which degree a participant believes that a certain linguistic expression characterises a probability value (Bisantz, Marsiglio and Munch 2005). Transferred to the current experiment, membership functions can describe to which degree an individual thinks that a particular stimulus belongs to a certain urgency level. This enables researchers to identify how confident users are in distinguishing several urgency levels (Bisantz, Marsiglio and Munch 2005).

To establish a basic familiarisation with the implemented variable instances, participants were initially shown all levels in the order as per designer intentions (see Table 6.3). This was required to establish the range of each variable, e. g. minimum and maximum size of the light strip. Following this introduction, the four levels were successively shown in randomised order. Succeeding each presentation of a certain variable instance, e. g. level 2 of the variable position, the participants were prompted to indicate on a scale from 0 to 10 to what degree they believe that the currently displayed instance corresponds to each of the four initially shown levels. Thus, for each of the urgency levels, participants had to provide four indications, i. e. one membership score (from 0 to 10) for each membership level. For example, if users were not able to distinguish if the currently displayed instance corresponds to urgency level 1 or to urgency level 2, they would indicate this with intermediate membership scores for membership levels 1 and 2 and low scores for the remaining membership levels. A variable that affords users to confidently distinguish several levels would receive high mean scores for the correct level and lower scores for the remaining, incorrect levels. The mean difference between the score for the correct level and that of the next highest level can serve as an indication for the accuracy of discrimination between the different uncertainty levels.

The logic of each variable for conveying uncertainties, i. e. research question 2 (see p. 159), and the perceived interruption, i. e. research question 3, were investigated using the following questionnaire items in conjunction with a 7-point Likert scale ranging from *strongly disagree* to *strongly agree*:

- The light variable logically communicates the capability of the system to safely perform the driving task.
- The light variable was annoying.

Apparatus Throughout the experiment, participants were seated in a driving simulator in automated mode to generate context (see Figure 6.6). The driving simulator was located in a soundproof laboratory with controlled lighting to ensure similar conditions for all participants. In line with the identified guidelines (see Subsection 6.1.3.1), the variables were conveyed using a light strip with a length of 50 cm and consisting of 77 LEDs that was positioned on top of the centre console. The light strip was controlled with 8-bit pulse width modulation (PWM) precision using an Arduino Mega 2560 Rev3 microprocessor and was driven by a 5 V (10 A) DC power supply. Participants interacted with a tablet computer attached to the centre console that displayed the experimental questions and recorded the participant responses.

Procedure Pre-recorded videos were used to explain the context of the experiment as well as the experimental procedure to participants. Following the video, participants filled in a consent form and were seated in the driving simulator, where they were prompted to adjust the seat to their usual driving position. For the duration of the study, participants were asked to focus at the centre of the tablet, which was highlighted by a small sticker. This was to ensure that the light strip would be perceived peripherally. Compliance with this requirement was monitored by the experimenter. Following a practice question, the

Figure 6.6: Experimental setup for the evaluation of peripheral light variables (Kunze, Summerskill et al. 2018b)

experiment started. After responding to all of the questions relating to the fuzzy membership functions, participants were prompted to indicate how logical and how interrupting they perceived each light variable. The complete scenario lasted approximately 30 min for each participant and was followed up by a debrief with the opportunity for any further remarks.

6.1.4.3 Results

The results of the experiment suggest a series of differences between the investigated light variables in regard to each of the specified research questions.

Figure 6.7 depicts the fuzzy membership functions for each light variable and urgency level. For extreme urgency levels (1 and 4), the membership functions show a monotonically decreasing (urgency level 1) and increasing trend respectively (urgency level 4). Thus, the mean sorting scores were highest for the membership level that corresponds with the urgency level. For intermediate urgency levels (levels 2 and 3), the trend is unimodal, whereby the membership level with the highest score corresponds with the urgency level for all variables

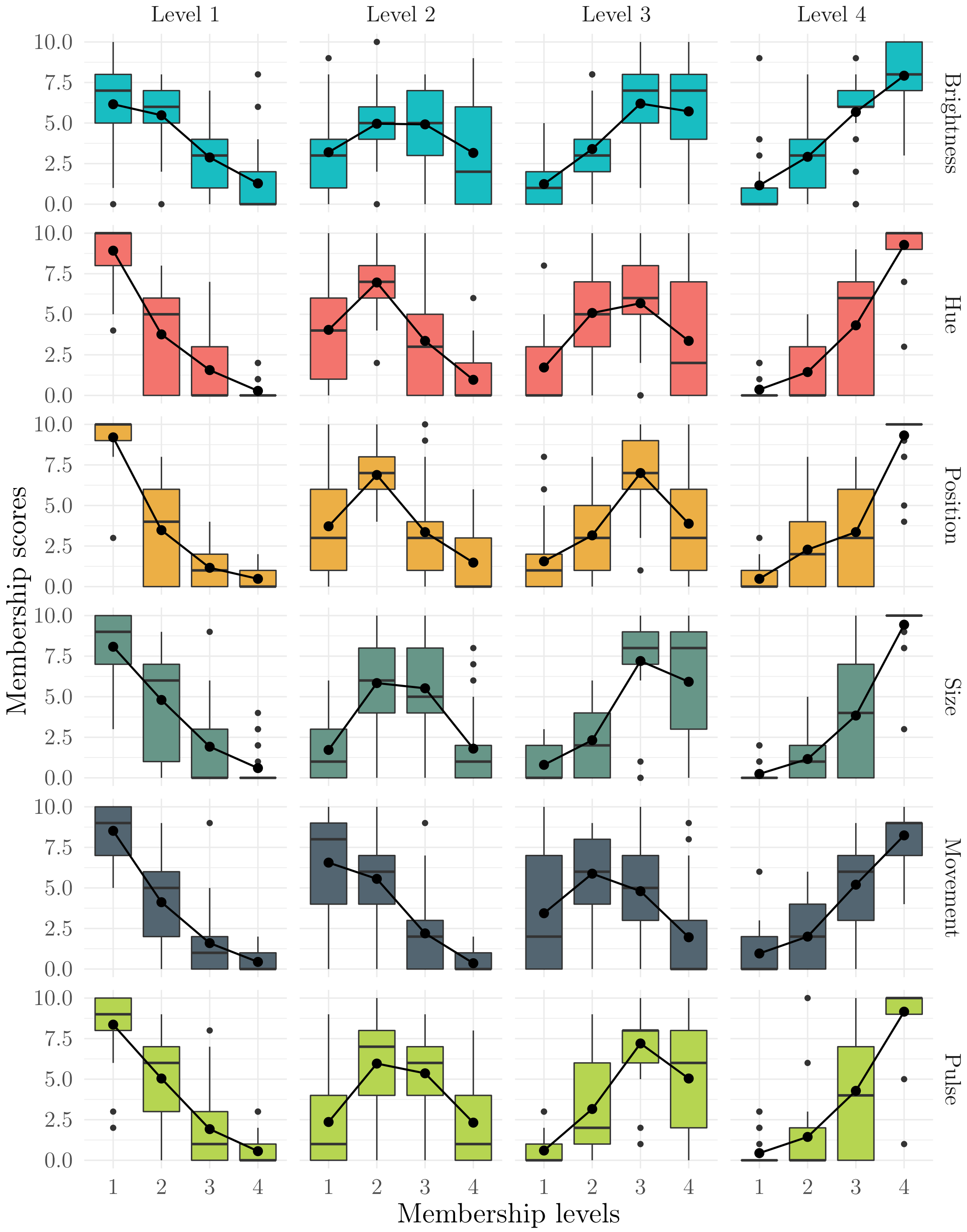


Figure 6.7: Membership functions for each light variable. Higher membership scores indicate an increased association between variable instance and membership level. The subcolumns depict the responses for each level. For instance, the leftmost box in the first row indicates the membership scores participants assigned to level 1 for the variable brightness (Kunze, Summerskill et al. 2018b)

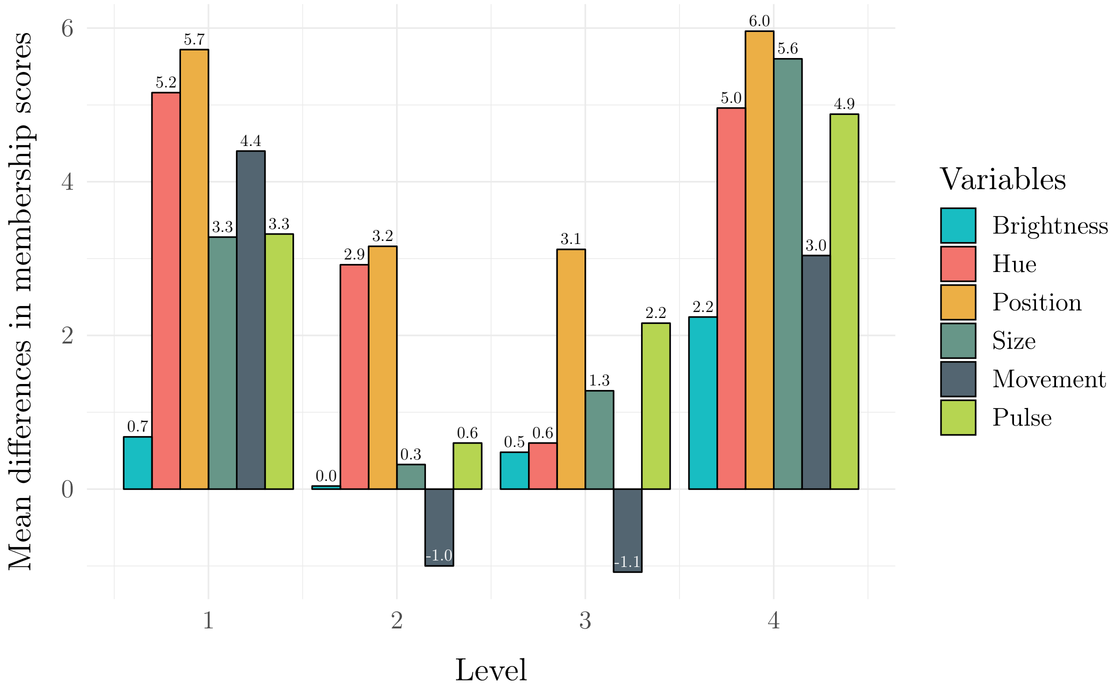


Figure 6.8: Mean differences in membership scores from correct membership level to next highest membership level. The higher the mean difference, the more confidently was the level distinguished from other levels. Negative mean differences indicate that an incorrect membership level received the highest mean score.

except movement. Figure 6.8 summarises the mean differences between the membership scores for the correct urgency level, e. g. score for membership level 1 regarding urgency level 1, and the highest membership scores for incorrect membership levels, e. g. score for membership level 2, 3, or 4 regarding urgency level 1. The higher the mean difference, the more confidently levels were distinguished from others. Negative values indicate that incorrect membership levels received higher scores than the correct membership level. Overall, extreme urgency levels were more readily matched with corresponding stimuli (see Table 6.3) than intermediate urgency levels, as indicated by higher mean differences. For each urgency level, the variable position received the highest mean differences, followed by hue, size, and pulse. Brightness and movement, with the exception of urgency level 1, received consistently lower scores than the remaining variables.

Table 6.5 summarises the responses to the questionnaire items for each variable, whereby the Likert-scale responses were quantified (1: *strongly disagree*, 7: *strongly agree*). Bartlett's test returned non-significant differences in the variances between variables for both items. A within-subjects one-way ANOVA on quantified Likert-scale scores was performed for each questionnaire item to assess the differences between means. The statistical model returned

Table 6.5: Descriptive statistics for logic and annoyance indications

Variable	**Logic**		**Annoyance**	
	M	SD	M	SD
Brightness	4.320	1.887	2.920	1.631
Hue	5.320	1.676	2.240	1.091
Position	3.520	2.044	2.840	1.519
Size	5.040	1.428	3.000	1.384
Movement	5.040	1.399	3.520	1.327
Pulse	5.240	1.234	4.600	1.756

Table 6.6: Results of Holm-corrected (Holm 1979) post-hoc t tests regarding the logic for conveying uncertainties

	Brightness	Hue	Position	Size	Movement
Hue	.356	-	-	-	-
Position	.775	.002**	-	-	-
Size	.976	1.000	.017*	-	-
Movement	.976	1.000	.017*	1.000	-
Pulse	.487	1.000	.004**	1.000	1.000

$^{*}p < .050$ $^{**}p < .005$ $^{***}p < .001$

Table 6.7: Results of Holm-corrected (Holm 1979) post-hoc t tests regarding perceived annoyance

	Brightness	Hue	Position	Size	Movement
Hue	.828	-	-	-	-
Position	1.000	.902	-	-	-
Size	1.000	.622	1.000	-	-
Movement	.902	.027*	.828	.902	-
Pulse	.001**	<.001***	<.001***	.002**	.102

$^{*}p < .050$ $^{**}p < .005$ $^{***}p < .001$

significant main effects for both logic, $F(5, 24) = 4.539, p < .001, \eta^2 = 0.136$ (medium), and annoyance, $F(5, 24) = 9.265, p < .001, \eta^2 = 0.207$ (large). Holm-corrected (Holm 1979) post-hoc t tests were performed to determine where the observed differences originate from. The results are summarised in Tables 6.6 and 6.7. As such, position was rated significantly lower in terms of logic for uncertainty communication than all other variables with the exception of brightness. The animation-based variables were rated as significantly more annoying than the remaining variables, whereby pulse received the overall highest mean score.

6.1.4.4 Discussion

The previously outlined experiment was conducted to examine which abstract variables are best suited for visually conveying uncertainties using a peripheral light strip.

The membership functions and mean differences indicate that participants were able to distinguish between each of the four urgency levels for all variables except movement. As indicated by mean differences in membership scores, particularly position afforded a clear discrimination between different levels, whereas changes in brightness and movement seem to be less easily distinguished. Thus, brightness and movement are rather unsuited for conveying several levels of information. The identified difficulties in detecting changes in brightness relative to, for instance, hue are in line with prior work (Davis et al. 2017). Differences between pulse and movement are potentially a consequence of increasing complexity. The distribution of rods and cones in the human retina (see Subsection 2.4.3.2) would suggest that the detection of differences in colour hue diminishes as a stimulus is moved towards the periphery. The results of this experiment, however, indicate that colour hues were, on average, correctly identified. This may be due to the positioning of the light strip in the near periphery of the driver (see Subsection 2.4.3.2).

Further, lower mean differences for intermediate levels suggest that middle values are not as readily distinguished. In line with this finding, Matviienko et al. (2015) found that the initial and final state of a colour is most significant for conveying information. This suggests that the number of intermediate values between end points should be reduced.

The responses regarding the logic of each light variable for uncertainty communication indicate that particularly a change in position is ill-suited for this application. This may be due to the directionless nature of uncertainties. Further, all variables but position increase in salience with higher urgency levels. In contrast, the variable hue was rated as most logic, which may be due to its familiarity (Faltalous et al. 2018; Ou et al. 2012). The animation-based variables, particularly pulse, were rated as significantly more annoying than other variables and are therefore better suited to convey highly urgent information

(Mueller et al. 2014). The higher annoyance of animation-based variables can be attributed to the relatively high sensitivity of the human retina for movements or changes in luminosity, i. e. pulses, in the periphery (see Subsection 2.4.3.2).

6.1.4.5 Limitations

The discussed results should only be considered within the confines of the following limitations. Validity is limited due to the use of different scales for each variable. Specifically, the mapping of one scale, for instance hue, to another scale, e. g. size, is unclear, thus impeding the general comparability between variables. However, the results are valid with regard to the initially outlined premisses. This includes the use of a light strip attached to the top of an automotive centre console, which limits the range of the variables size and position. Further, the use of learned representations regarding colour condition the variable hue. Additionally, accessibility aspects such as the prevention of photosensitive epilepsy limit the frequency of animations.

Moreover, the experiment did not investigate how cultural influences affect the results. For instance, colour hues might have a different meaning in other cultures (Heimgärtner, Solanki and Windl 2017). Further, the study setting does not allow a detailed assessment of how peripheral vision is affected by each of the variables because of variances in seating position and, consequently, differences in the position of the light strip within the FOV. Rather, the study aimed at representing a realistic driving situation in which the driver would need to perceive the information.

6.1.4.6 Conclusion

The results of the experiment provide a series of guidelines that supplement those presented in Subsection 6.1.3.1. As such, the variables brightness and movement should be avoided when conveying several levels of information using peripheral lights. Different instances of both variables were not as readily matched with the corresponding urgency level as the remaining variables. While differences in position proved to be easiest to distinguish, the variable was rated significantly lower in terms of logic for uncertainty communication than all other variables except for brightness and should therefore be avoided for this application as well. The remaining variables, hue, size, and pulse, were all found to be somewhat suitable for uncertainty communication, whereby pulse was perceived as most annoying and should therefore be reserved for highly critical information.

6.1.5 Evaluation of Vibro-Tactile Variables

Vibro-tactile feedback can supplement the information conveyed by peripheral light strips to imply higher urgency levels. As highlighted by effects such as inattentional blindness (see Subsection 2.4.3.1), relying solely on visual feedback may lead to missed changes, prompting the need for stimuli that are less likely to be missed, such as vibro-tactile cues. In contrast to visual feedback which can be conveyed without necessarily distracting users, vibration leads to attentional spikes that impede continual vibro-tactile feedback (see Subsection 6.1.1.4). Rather, vibro-tactile stimuli can be employed to notify users of significant transitions from one urgency level to another.

In Subsection 6.1.3.2, four vibro-tactile characteristics were derived from existing research that are generally eligible for indicating transitions: vibration amplitude, position, moving pattern, and rhythm. It remains to be addressed which characteristics are best suited for conveying uncertainty transitions, whereby both increases and decreases are of interest. On one hand, users may be engaged in an NDRT and require vibro-tactile feedback to be notified of increases in uncertainty. On the other hand, users may have been prompted to focus on the FRD as a consequence of increased uncertainty but may fail to notice that the uncertainty has decreased if only visual feedback is provided. Thereby, different directions or instances of the vibro-tactile feedback have to be taken into account. For instance, amplitude can be increased or decreased, movement can be up or down. Consequently, the following research questions are formulated:

1. Which vibro-tactile variable instances clearly convey an increase in uncertainty?
2. Which vibro-tactile variable instances clearly convey a decrease in uncertainty?

As a prerequisite for addressing these research questions, it remains to be determined how the vibro-tactile variables are to be specified for this evaluation.

6.1.5.1 Selection and Specification of Vibro-Tactile Variables

The selection of basic vibration patterns is grounded in the variables identified by Ji, Lee and Hwang (2011). As depicted in Figure 6.9, changes in amplitude, position, movement, and rhythm were evaluated, whereby the number and positioning of the actuators is based on the seat layout presented in Chang, Hwang and Ji (2011) (see Figure 6.5). The overall number of actuators was similar for each variable for reasons of comparability. For each of the variables, two opposing patterns were created (see Figure 6.9). For amplitude pattern A, the vibration intensity was gradually increased from 0 % to G_{max}, whereby G_{max} was specified in agreement with previous work (see Table 6.2 and Subsection 6.1.5.2). For pattern B, the change in amplitude was reversed. To accommodate the requirement of

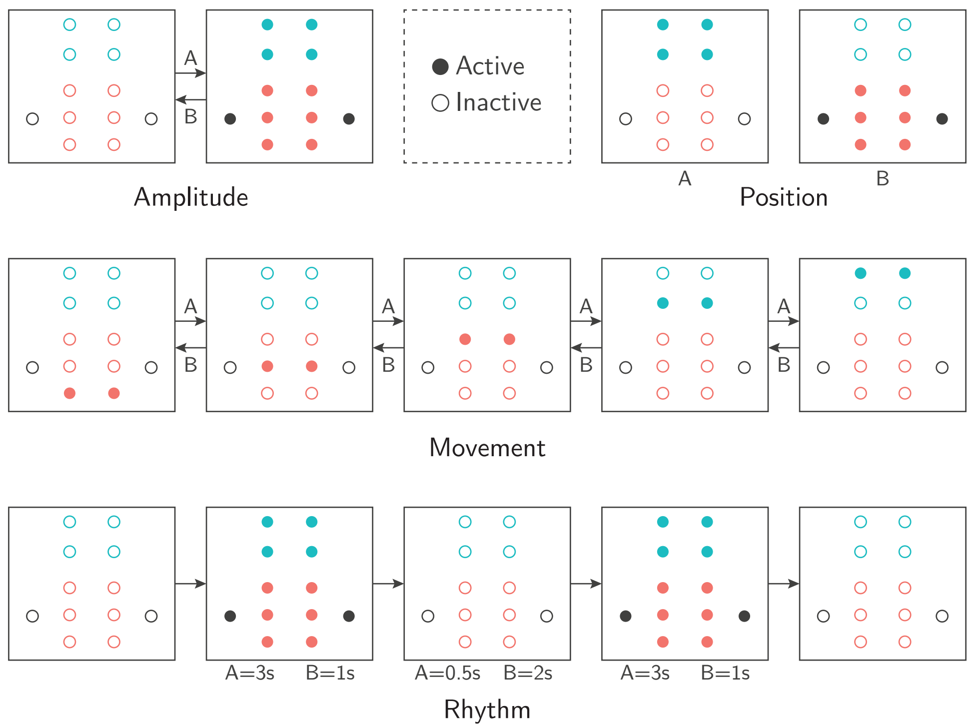


Figure 6.9: Vibro-tactile variable instances derived from Ji, Lee and Hwang (2011) and employed in the outlined experiment (Kunze, Summerskill et al. 2018c)

stimuli with a duration of at least 10 ms (see Subsection 6.1.3.2), a delay of 60 ms was implemented between incremental changes in amplitude. The maximum amplitude was used for all remaining variables to promote internal validity. Position was varied between backrest (pattern A) and seat pan (pattern B). As the backrest is positioned higher than the seat pan, it was hypothesised that pattern A would be rated more favourably for conveying increases while pattern B would be preferred for decreases . Based on the same assumption, movement was implemented as a wavelike upwards (pattern A) and downwards movement (pattern B) (Schwalk, Kalogerakis and Maier 2015). Following the specifications proposed by Ji, Lee and Hwang (2011), rhythm was varied between 3 s impulses separated by 0.5 s gaps (pattern A) and 1 s impulses separated by by 2 s gaps (pattern B).

6.1.5.2 Method

The research questions (see p. 159) were investigated using an identical laboratory setting and involving the same participants as the experiment outlined in Subsection 6.1.4.2. Ethical approval was granted by the Ethics Approvals Sub-Committee of Loughborough University.

Experimental Design A within-subjects design was employed to address the research questions, whereby the order of variables and patterns was randomised to prevent order effects. Questionnaires with 7-point Likert scales were used to evaluate how clearly changes in uncertainty are being conveyed. For each pattern, participants were prompted to state their agreement with the following questions:

- The vibration pattern clearly communicates that the vehicle is struggling more than before.
- The vibration pattern clearly communicates that the vehicle is struggling less than before.

The term uncertainty was removed as a consequence of ambiguities identified during pilot testing. The concept of struggling was found to be more relatable for novice users of automated driving systems.

Apparatus The experimental setting outlined in Subsection 6.1.4.2 was replicated in this study (see Figure 6.6). The peripheral light strip remained deactivated throughout the experiment. As specified in the guidelines (see Subsection 6.1.3.2), eccentric rotating mass vibration motors were placed into a vehicle seat, whereby the layout shown in Figure 6.5 was used as a template. In agreement with the findings of Ji, Lee and Hwang (2011), the actuators were placed more than 9 cm apart from each other beneath a fabric layer consisting of 1.5 mm leather and 5 mm foam (see Table 6.2). The actuators were thereby mounted directly in cut-out, sealed pockets of the foam layer. Following the mounting of all actuators, the seat was returned to its original state. In reference to prior research (see Table 6.2), the maximum normative amplitude was set to approximately 2.65 G (70 Hz, 1.6 V, see Figure A.7).

The voltage of the motors was controlled using an Arduino Mega 2560 Rev3 microprocessor. To accommodate the current draw of the actuators, metal-oxide semiconductor field-effect transistors (MOSFETs) were used. Acting like a switch, the transistor is controlled by the low current originating from the output pin of the microcontroller and connects to a DC supply with a sufficient current to power the vibration motors. Figure A.8 shows the circuit used to power and control the vibration motors. A pull-down resistor (100 kΩ) was added to lower quiescent current and increase system stability. A purpose-built printed circuit board (PCB) was designed to electrically connect the components.

Procedure The experiment was conducted subsequent to the previously described study (see Subsection 6.1.4.2). Following an instruction about the purpose and procedure of the study, the different vibration patterns (see Figure 6.9) were presented to participants in a randomised order. After three consecutive presentations of each vibration pattern, the vibration motors were deactivated and the participants were prompted to indicate their agreement with the previously described statements. After completing the responses for all variables, the participants were debriefed and were given the opportunity for any further remarks.

6.1.5.3 Results

Table 6.8 and Figure 6.10 summarise the results of the experiment for each variable pattern and uncertainty transition direction. Overall, amplitude pattern A received the highest agreement rating for indicating increases in uncertainty ($M = 5.680, SD = 1.145$), followed by rhythm pattern A ($M = 5.360, SD = 1.440$). The same patterns also received the lowest ratings for decreases ($M = 2.200$).

A linear mixed effects model was performed to assess the impact of uncertainty transition direction (increasing; decreasing) and variable pattern on agreement scores. There was a significant main effect of uncertainty transition direction, $t(391) = 12.936, p < .001$. This indicates that the agreement scores were higher for the statement indicating increases in uncertainty.

Paired t tests were performed to assess the differences between the opposing statements for each variable pattern (see Table 6.9). To unambiguously indicate increases or decreases in uncertainty, the differences between the agreement scores should be maximised. With the exception of movement and rhythm patterns B, the differences were significant, whereby the largest mean differences were recorded for amplitude pattern A ($MD = 3.480$) and rhythm pattern A ($MD = 3.160$).

6.1.5.4 Discussion

The results indicate a series of insights for using vibro-tactile variables for uncertainty communication in an automotive context. As vibro-tactile feedback is not as ubiquitous in an automotive context as visual or auditory cues, the tactile stimuli should be self-explanatory (Erp 2002).

Irrespective of the vibro-tactile pattern, the stimuli conveyed increases in uncertainty, as indicated by a significant main effect of the uncertainty transition direction. Thus, vibro-tactile stimuli are not suitable for conveying reductions in uncertainty, i. e. positive information. Rather, the vibration input should be limited to conveying increasing urgency, i. e. information indicating imminent critical situations or hazards. This is in line with other approaches in the automotive domain employing vibro-tactile feedback to convey critical information. For instance, vibro-tactile feedback has been used to indicate TORs

Table 6.8: Descriptive statistics for the agreement with each statement regarding increasing/decreasing uncertainty (Kunze, Summerskill et al. 2018c)

		Uncertainty transition			
		Increasing		Decreasing	
Variable	**Pattern**	*Mean*	*SD*	*Mean*	*SD*
Amplitude	A	5.680	1.145	2.200	1.155
	B	4.880	1.900	3.040	1.859
Movement	A	5.080	1.706	2.640	1.630
	B	4.640	1.934	3.240	2.026
Position	A	4.800	1.190	2.640	1.319
	B	4.640	1.524	3.280	1.370
Rhythm	A	5.360	1.440	2.200	1.041
	B	4.080	1.631	3.320	1.574

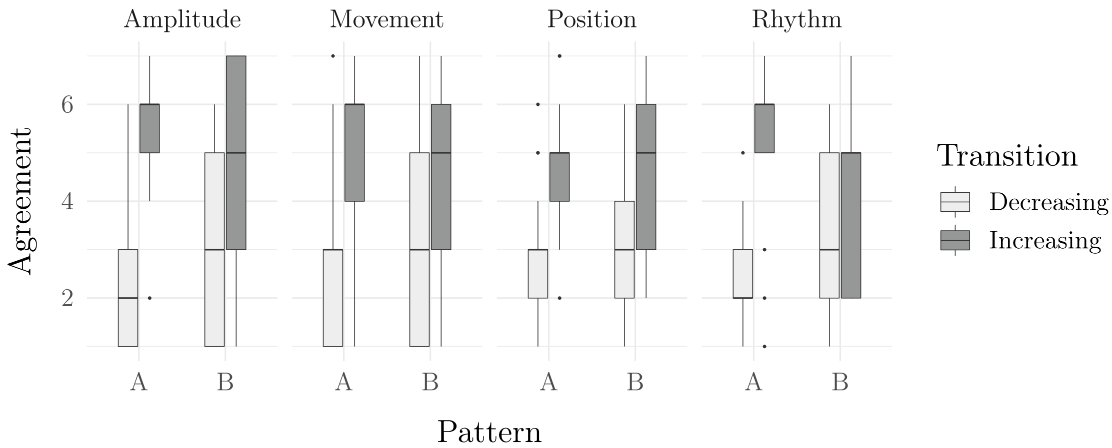


Figure 6.10: Box plots summarising agreement with statements regarding increases and decreases in uncertainty for each vibro-tactile pattern (Kunze, Summerskill et al. 2018c)

(Schwalk, Kalogerakis and Maier 2015) or speed warnings (Chang, Hwang and Ji 2011) and is currently employed in production vehicles for communicating collision alerts (GMC 2018).

The preference of participants for longer vibration durations separated by shorter intervibration intervals, as indicated by rhythm patterns A and B, is in accordance with other work (Ji, Lee and Hwang 2011). Variations in movement and position scored similarly high for conveying increases in uncertainty. As such, the initially formulated hypothesis that *moving the vibration upwards* is intuitive for conveying increases (see p. 170) cannot be confirmed.

Table 6.9: Results of paired t tests between responses regarding direction (increasing/decreasing) (Kunze, Summerskill et al. 2018c)

Variable	Pattern	Increasing/Decreasing	MD	Cohen's d
Amplitude	A	$t(24) = 8.046^{***}$	3.480	3.027
	B	$t(24) = 2.563^{*}$	1.840	0.979
Movement	A	$t(24) = 3.794^{***}$	2.440	1.463
	B	$t(24) = 1.860$	1.400	0.707
Position	A	$t(24) = 4.726^{***}$	2.160	1.719
	B	$t(24) = 2.542^{*}$	1.360	0.938
Rhythm	A	$t(24) = 6.969^{***}$	3.160	2.515
	B	$t(24) = 1.277$	0.760	0.474

$^{*}p < .050$ $^{**}p < .005$ $^{***}p < .001$

The patterns amplitude A and rhythm A received the highest mean scores for conveying increases in uncertainty and the largest mean differences between the opposing questionnaire items. This suggests that both can be unambiguously interpreted as signifiers for increases in uncertainty (as opposed to decreases).

6.1.5.5 Limitations

The findings are confined by several limitations. While the experimental setting allowed for the controlled testing of several variable patterns, it did not ensure that the patterns can be understood in a dynamic driving environment where they are to be processed in conjunction with other information. Consequently, additional investigations are needed that more closely replicate the specific usage context. Further, the impact of age, gender, or weight on experiment parameters such as maximum amplitude was not considered, although these might have an impact (Ji, Lee and Hwang 2011). Practical implementations of vibro-tactile seat interfaces should calibrate the maximum vibration intensity depending on the aforementioned parameters.

6.1.5.6 Conclusion

To the knowledge of the author, the outlined results show the first implications for communicating uncertainty changes using vibro-tactile stimuli.

The conducted experiment affords the deduction of the following guidelines. First, vibro-tactile stimuli should only be used to convey increases in uncertainty. Second, particularly increases in amplitude and rhythms consisting of long vibrations (3 s) separated by short intervals (0.5 s) are best suited for conveying these increases.

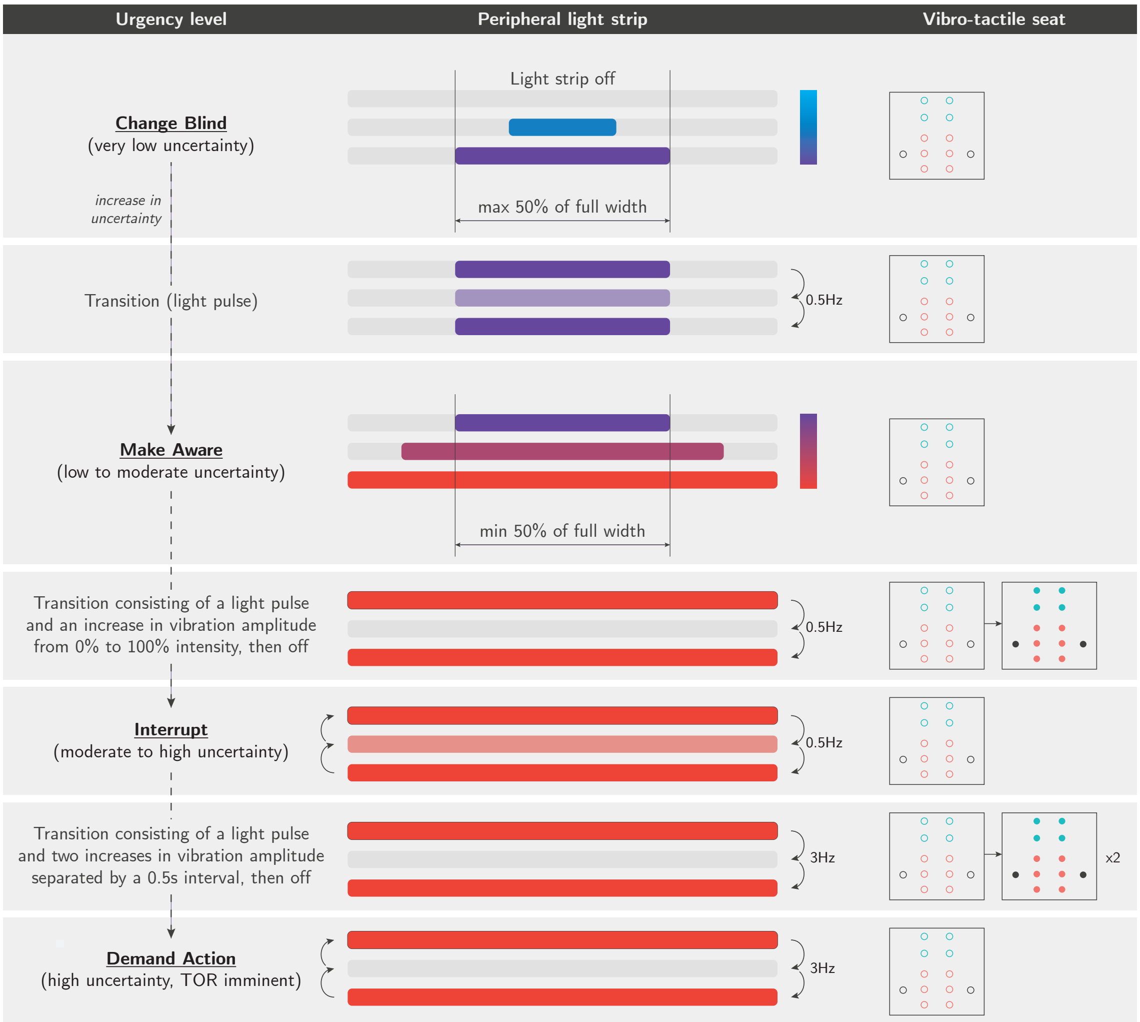


Figure 6.11: First display design iteration

6.2 Concept Iteration

The second phase of the process specified in Figure 5.8 aims at fulfilling the requirements stated in the PADE framework (see Figure 5.5) within the previously established guidelines. This section outlines the iterative development process of a peripheral awareness display for uncertainty communication. Starting with a first iteration derived from the guidelines presented in Section 6.1, the designs are optimised based on two separate evaluations. In a first step, five usability experts completed an inspection-based, heuristic evaluation. The updated design was then subjected to 24 novice participants in a user-based evaluation. The final design serves as a basis for investigating the implications of peripheral uncertainty communication discussed in the subsequent chapter.

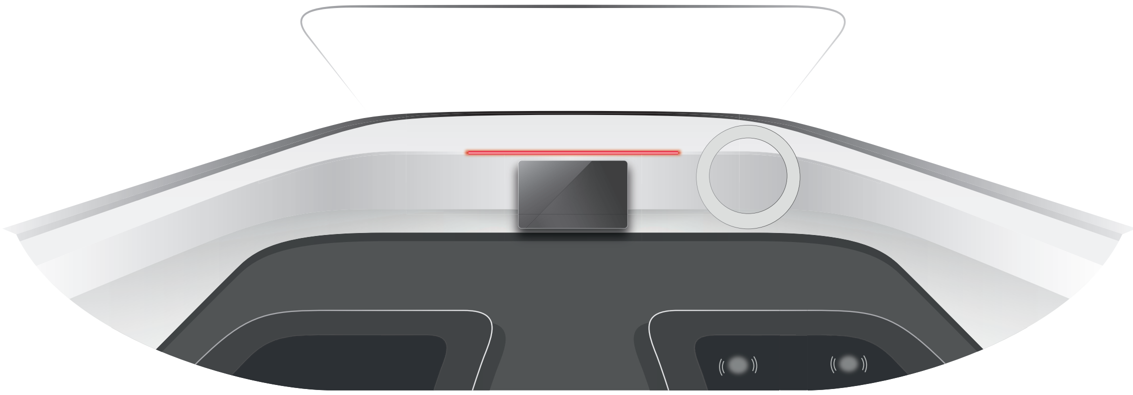


Figure 6.12: Implementation of the display design in an automotive cockpit

6.2.1 First Iteration of a Peripheral Uncertainty Display

Figure 6.11 summarises the interaction design of the peripheral uncertainty display in its first iteration. Initially, the four urgency levels proposed by Matthews, Dey et al. (2004) and implemented in the PADE framework (see Figure 5.5) were used as a basis. Following the previously established guidelines and the specified usage context (see Section 6.1), the design takes a multimodal approach, addressing both the peripheral visual and the tactile channel. As depicted in Figure 6.12, the feedback was implemented using a peripheral light strip attached to the top of the centre console and vibration motors mounted to the vehicle seat. The cockpit layout is adopted from the experiments shown in sections 6.1.4 and 6.1.5.

The uncertainty information was redundantly coded for the peripheral light strip to make it accessible to users suffering from colour blindness (see Table 6.1) and harness the benefits of redundancy for visual processing (see Subsection 2.4.3.1). As discussed in Subsection 6.1.4.6, the variables hue and size are both perceived as logical for uncertainty communication and users can readily distinguish several levels of information based on either. Hue was varied using the temperature metaphor (see Subsection 6.1.3.1), whereas the size of the light was increased with higher levels of urgency. Animation-based variables were reserved for the highest levels of urgency, i. e. *interrupt* and *demand action*, as they were perceived as most interrupting (see Subsection 6.1.4.4). The pulse frequency was increased to distinguish the levels, whereby its value was specified according to the extreme levels 1 and 4 (see Table 6.3) in order to ensure their confident identification. Thereby, the levels can be considered as distinct visualisations, with gradual changes only being integrated for aesthetic purposes and to indicate transitions. The colour for uncertainty values corresponding to the *make aware* urgency level, for instance, does not gradually change between purple and red depending on minor uncertainty changes, but remains at a red colour as long as the uncertainty is low to moderate.

Transitions to higher uncertainty levels were supplemented by vibro-tactile feedback. Following the evaluation of vibration patterns in Subsection 6.1.5, increases in amplitude and a rhythm consisting of long vibrations (3 s) separated by short intervals (0.5 s) were used to convey the transitions.

To account for sudden increases in uncertainty, the display can skip lower urgency levels and immediately convey the *demand action* pattern. Thereby, *demand action* does not convey a TOR and thus demands a takeover as user action, but rather requires the user to gaze towards the FRD.

6.2.2 Heuristic Evaluation

In reference to the development process shown in Figure 5.8, the first iteration of a peripheral uncertainty display was evaluated using an inspection-based heuristic evaluation in order to identify existing usability issues (see Subsection A.3.5). This involves the recruitment of expert evaluators who assess the compliance of the concept display with a set of supplied heuristics, i. e. usability guidelines.

6.2.2.1 Recruitment of Evaluators

Several factors influence the number of evaluators, including the anticipated problem frequency, their impact, and their visibility (Woolrych and Cockton 2001). The current display concept can be considered low in both complexity and diversity as it conveys information along a single dimension without any degrees of freedom in terms of personalisation. Consequently, five evaluators were recruited for the heuristic evaluation (see Subsection 5.2.3.3). To be eligible, they had to have at least five years of experience designing or researching on human-machine interfaces. The research group for user centred design of the School of Design and Creative Arts at Loughborough University served as a recruitment pool.

6.2.2.2 Procedure and Identified Issues

The evaluators were provided with images and a worded description of the peripheral uncertainty display (Nielsen and Molich 1990) accompanied by an animation that simulates how the peripheral light strip changes with increasing uncertainty. Further, they were given a description of the heuristics. They were instructed to assess the interface in terms of its compliance with each heuristic. Identified issues were to be sorted in reference to the heuristics and had to be supplemented by a severity score ranging from 1 to 5. Scores ranging from 1 to 2 indicated minor issues that posed irritations but no significant barriers for usage. A score of 3 suggested somewhat critical, intermediate issues. The remaining

Table 6.10: Results of the heuristic evaluation involving five expert evaluators

Heuristic	E1	E2	E3	E4	E5
Sufficient information design	•1	•2	-	•3	-
Consistent and intuitive mapping	•5	•4	•3	•4	-
Match between system and real world	-	•3	-	-	-
Visibility of state	-	-	•3	•3	-
Aesthetic and pleasing design	•3	-	-	-	-
Useful and relevant information	-	•3	-	•3	-
Visibility of system status	-	•3	•4	-	•5
Peripherality' of display	-	-	-	-	-
Flexibility and efficiency of use	-	-	•3	-	-
Help users recognise, diagnose, and recover from errors	•4	-	•3	-	•2

E1-5: expert evaluators, •: issue discovered for heuristic, 1-5: severity rating

scores indicated severe issues that might critically discourage or hinder the use of the interface.

Table 6.10 summarises the issues that were identified regarding each heuristic. Thereby, not all heuristics applied to the interface design. As there was no input from the user and no means of personalisation, the heuristics *user control and freedom*, *easy transition to more in-depth information*, and *error prevention* did not apply and are excluded from the table. In total, six severe issues (severity ≥ 4) were reported. As such, evaluators stated that it may be difficult to remember what the colours refer to (severity: 5, count: 1) and what the changes in state mean (4, 1). Addressing these issues, one evaluator suggested to provide a verbal introduction to explain the different states (3, 1). Further, it was reported that a red light with full size, currently used to convey the *make aware* level, indicates critical situations and should be reserved for *interrupt* and *demand action* (4, 1). Also, vibration should be reserved for the highest level and only when an action is required (5, 1). Additionally, evaluators were missing an indication of the working state (4, 1) and a vibration to indicate errors with the display (4, 1). Four times it was stated that the *change blind* level is unnecessary because the information is always somewhat relevant (2-3, 4). Rather, the display should remain off for very low levels to reduce cognitive workload (3, 1). Further, it was expressed that the transitions incorporating pulses with 50 % intensity, e. g. from *change blind* to *make aware*, are likely not perceptible in bright lighting environments (3, 2). Regarding the heuristic *aesthetic and pleasing design*, it was reported that the display must be implemented in a way that matches the overall vehicle interior (3, 1). Additionally, an error of the display should be accompanied by a verbal message precisely indicating the problem (2-3, 2). Two evaluators reported that the system does not necessarily match user expectations as it does not use the traffic light metaphor (3, 2). Finally, auditory feedback may be provided to supplement the information (1, 1).

Figure 6.13: Second display design iteration

6.2.2.3 Second Iteration of a Peripheral Uncertainty Display

Figure 6.13 depicts the amended interaction design under consideration of the identified issues during the heuristic evaluation. Following the feedback by expert evaluators, it incorporates a reduction in levels from four to three, omitting the *change blind* level. Very low uncertainty values are now represented by a deactivated light strip to reduce cognitive workload by removing one of the monitoring tasks. Further, the transitions are now implemented with complete pulses and the fully red light is reserved for the levels *interrupt* and *demand action*. LEDs for the operational state were added, whereby transitions to display errors are indicated by a vibration with the opposite direction of indicating an increase in urgency i. e. reduction in amplitude. A change in favour of the traffic light metaphor was not incorporated due to the arguments presented in Section 6.1. As such, the traffic light metaphor is more susceptible to accessibility problems and does not exhibit the intended active/passive connotation regarding monitoring behaviour. As suggested by one of the reviewers, the meaning of the different states must be explained in an introduction or tutorial. Further, redundant verbal information can be provided in other vehicle displays such as a HUD or in the instrument cluster.

6.2.3 User-Based Evaluation with PADE

As outlined in the development process depicted in Figure 5.8, the second iteration cycle involves user-based testing under consideration of the intended usage context. For this, the PADE framework was applied in a driving simulator experiment (see Figure 5.5).

6.2.3.1 Method

The specified usage context (see Table 6.1) was replicated in a driving simulator to evaluate the designed peripheral uncertainty display against the requirements defined by the PADE framework (see Figure 5.5). Specifically, this experiment focused on responses to the PADE questionnaire to derive an overall score for the user experience afforded by the display (see Table 5.5). Thereby, the digital display employed in the experiment outlined in Chapter 4 served as a baseline. Table 6.11 summarises how the uncertainty levels were mapped to each other.

The experiment aimed at identifying (a) differences in user experience relative to the digital display and (b) potential shortcomings that serve as starting points for further optimisation. Additionally, it aimed at assessing the internal consistency of the questionnaire scales to determine and improve their reliability.

Aspects relating to human-automation interaction, such as trust or takeover performance, were not evaluated in this study. Rather, the experiment focused on the interaction design of the display itself. The implications of peripheral uncertainty communication are investigated in the subsequent chapter.

Table 6.11: Uncertainty levels as mapped to each display type

Uncertainty level		Peripheral	Digital
0	Very low	Display off	50 bpm
1	Low to moderate	Make Aware	80 bpm
2	Moderate to high	Interrupt	110 bpm
3	Very high	Demand Action	140 bpm

Participants A total of 24 participants (14 female) with a mean age of 27.417 years ($SD = 6.795$, range: $21 - 55$) volunteered to take part in the experiment. Participants held a driving licence for an average time of 9.438 years ($SD = 6.456$, range: $3 - 37$) and stated to drive an average of 5686.125 miles per year ($SD = 4126.792$). Participants that took part in the experiment described in Chapter 4 and evaluators that completed the heuristic evaluation were excluded from the current study to prevent an impact of prior familiarisation with the interface design of either display.

Experimental Design To facilitate the comparability of the two displays types, a within-subjects design was employed. Each participant experienced both displays in succession, whereby the order in which the display types were presented was balanced.

According to the specified usage context (see Table 6.1), participants were required to focus on a visually demanding NDRT similar to the visual search task used in Chapter 4.

Dependent Variables Responses to the PADE questionnaire (see Table 5.5) served as a key indicator for user experience. Variables not covered in the questionnaire, i. e. effectiveness, efficiency, and reaction, were assessed using behavioural data. The uncertainty display can be considered effective if it results in an increased attentional focus on the FRD with higher uncertainty levels (see Table 6.11). The efficiency of monitoring the uncertainty display can be derived from the NDRT solving rate during the lowest uncertainty level (very low, see Table 6.11) for which no attention on the FRD is required. Reaction can be measured as the accurate and fast response to a display stimulus (see Subsection 5.2.3.1). Specifically, the time gap between an increase in uncertainty and users' change in gaze towards the FRD served as a numeric indicator. Motor expressions were not evaluated as these are likely more affected by the primary task (Thüring and Mahlke 2007). The internal consistency of the individual questionnaire scales was assessed using Cronbach's alpha (see Section 5.3).

Apparatus The conceptual interaction design displayed in Figure 6.13 was implemented in a driving simulator cockpit (see Figure 6.6) to achieve a high-fidelity prototype. Similar to the apparatus used in the experiments described in subsections 6.1.4 and 6.1.5, both the light strip and the eccentric rotating mass vibration motors were controlled using an Arduino Mega Rev3 2560 microcontroller. The Tobii Pro Glasses 2 wearable eye tracker was used to record operator gaze behaviour at a sampling frequency of 100 Hz. An iPad Air 2 mounted on the centre console was used to display the visual search task (see Figure 3.5).

Procedure and Scenario Prerecorded videos were used to brief the participants about the purpose and procedure of the experiment. After filling in a consent form and demographic survey, each participant completed a practice scenario with a total duration of approximately 15 min. Following the introduction, the first display type was explained to the participant and the experimental scenario was started. In agreement with the four urgency levels of the PADE framework, a total of four uncertainty levels were implemented in the scenario: *very low*, *low to moderate*, *intermediate to high*, and *very high* (see Table 6.11, Figures 7.2 and 7.3). Thus, participants experienced all urgency levels of the display. Immediately after finishing the 20 min scenario, participants completed the PADE questionnaire on the tablet computer. The remaining display was subsequently evaluated in a similar manner. The experiment was concluded by a debriefing session.

Table 6.12: Descriptive statistics of the UX scores for each questionnaire section and display type

Section	Peripheral		Digital	
	M	SD	M	SD
Usability	6.306	0.840	5.208	1.633
Attention	6.175	1.140	4.754	1.763
Non-Instrumental	4.259	1.852	3.736	1.786
Emotional	4.816	1.583	3.830	1.730
Overall appraisal	6.000	1.126	4.389	1.716

6.2.3.2 Results

Table 6.12 and Figure 6.14 summarise the results of the PADE questionnaire. The specification of the design dimensions information criticality and aesthetic emphasis determines the weighting of each section (see Subsection 5.2.4.2). As discussed in Subsection 6.1.2, information criticality was specified with 5/5 whereas the aesthetic emphasis was set to 4/5, leading to a relative weighting of 0.278 for instrumental qualities and 0.222 for non-instrumental qualities (see equations 5.1 and 5.2). Inserting the values shown in Table 6.12 into Equation 5.3 results in the following aggregated UX scores:

$$
\begin{aligned}
\mathrm{X}_{\text{Peripheral}} &= \frac{6.306 + 6.175}{2} * 0.278 + 4.259 * 0.222 + (4.816 + 6.000) * 0.250 = 5.384 \\
\mathrm{X}_{\text{Digital}} &= \frac{5.208 + 4.754}{2} * 0.278 + 3.736 * 0.222 + (3.830 + 4.389) * 0.250 = 4.269
\end{aligned}
$$

Both display types received, on average, scores above the centre of the Likert scale (4). The peripheral uncertainty display was rated higher for each individual section (see Table 6.12), whereby the mean difference was largest for the overall appraisal, $MD = 1.611$, followed by the attentional subscale, $MD = 1.421$. Regarding individual items, the largest mean

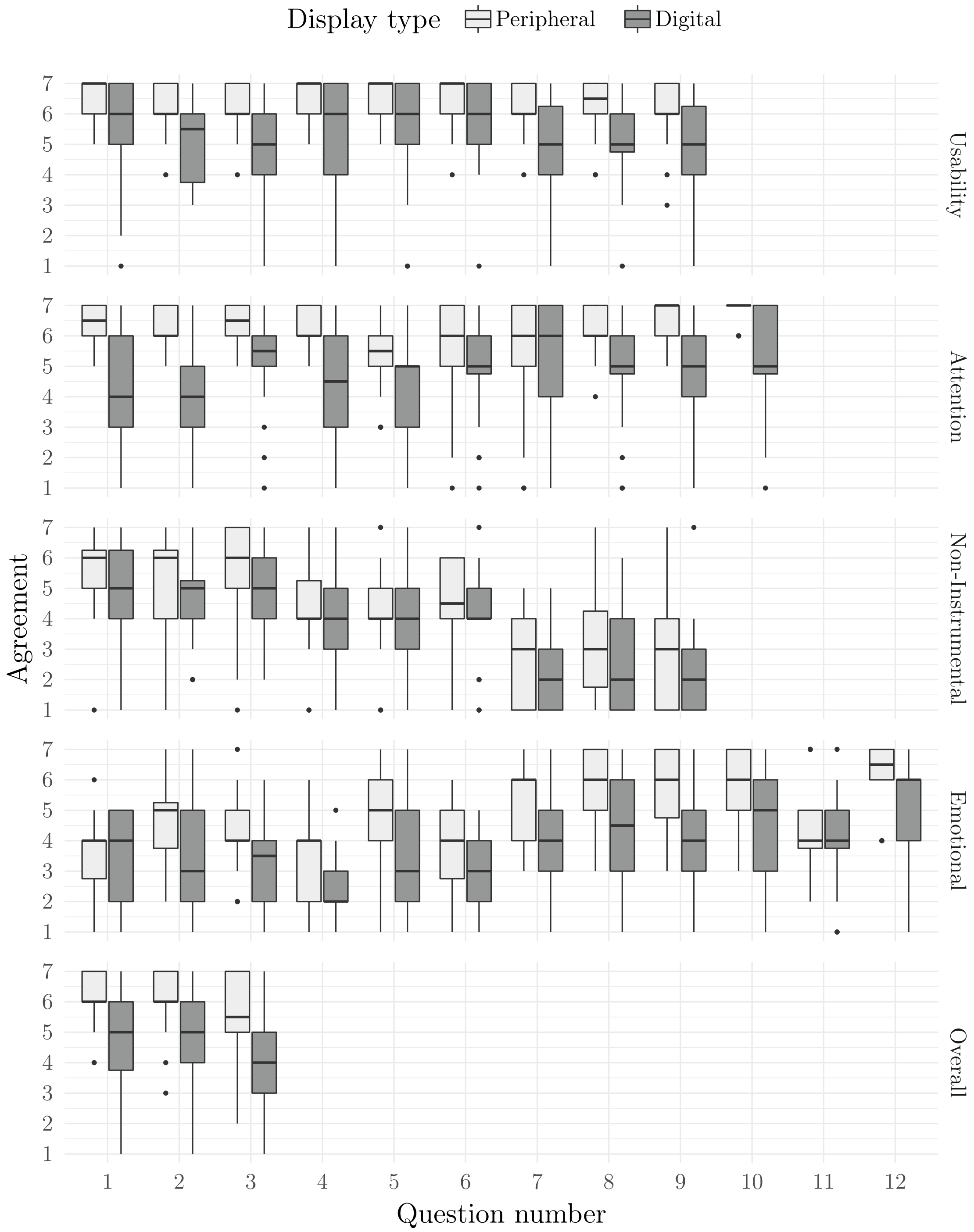


Figure 6.14: Box plots summarising agreement with statements regarding each PADE questionnaire item (question numbers are ordered as specified in Table 5.5, bullets depict outliers). Responses to questions relating to the scale negative emotions were reversed.

difference was observed for the first ($MD = 2.125$) and the second ($MD = 2.333$) item of the attention subscale. The lowest scores for both the peripheral uncertainty display ($M = 2.986$, $SD = 1.691$) and the digital display ($M = 2.375$, $SD = 1.438$) were received for the commitment scale (non-instrumental questions 7 to 9, see Figure 6.14).

For both displays, an increased monitoring ratio concerning the FRD could be observed with higher uncertainty levels. As such, participants interacting with the digital display monitored the FRD with a ratio of 0.055 ($SD = 0.049$) for very low uncertainties and 0.318 ($SD = 0.119$) for very high uncertainties. When using the peripheral awareness display, the ratio was 0.045 ($SD = 0.053$) for the lowest uncertainty level and 0.451 ($SD = 0.165$) for the highest level.

In situations characterised by very low uncertainties, participants achieved an NDRT solving rate of $29.868\,1/\min$ ($SD = 6.297\,1/\min$) when using the digital display and $30.855\,1/\min$ ($SD = 5.205\,1/\min$) when interacting with the peripheral awareness display. Results of a paired t test indicate that the difference in means is not statistically significant at the .05 level.

When the display changed from very low to increased uncertainty levels (see Figure 7.2), participants took on average $7.813\,\mathrm{s}$ ($SD = 8.194\,\mathrm{s}$) to focus on the FRD when using the digital display as opposed to $2.609\,\mathrm{s}$ ($SD = 2.688\,\mathrm{s}$) when interacting with the peripheral awareness display, $t(19) = 3.074, p = .006, d = 0.854$ (large).

Table 6.13 summarises the values for Cronbach's alpha regarding each questionnaire scale. With the exception of the scale comprehensibility, all values are above the critical threshold of 0.70 (see Section 5.3). Removing the item characterised by the least correlation with the remaining items of the scale, *I was able to understand the information just by glancing at it*, increased the value for Cronbach's alpha to 0.855 for the peripheral uncertainty display and to 0.731 for the digital display.

Table 6.13: Cronbach's alpha for each PADE questionnaire scale

Scale	**Cronbach's alpha**	
	Peripheral	Digital
Learnability	0.877	0.808
Comprehensibility	0.760	0.691
Usefulness	0.878	0.930
Attentional	0.787	0.861
Positive emotions	0.924	0.854
Negative emotions	0.825	0.790
Overall appraisal	0.911	0.915

6.2.3.3 Discussion

Overall, both display types were well-received by participants, as indicated by aggregated UX scores above the centre of the scale. The aggregated as well as individual UX scores, however, suggest that the peripheral awareness display was preferred over the digital display.

Particularly the attentional aspects may have contributed to this variation, as indicated by large mean differences in the corresponding questionnaire category and significantly different reaction times. As such, changes in urgency (attentional item 1, see Table 5.5 and Figure 6.14) were more difficult to perceive using the digital display and participants were less able to smoothly shift their attention between the NDRT and the uncertainty display (attentional item 2). Several reasons may have contributed to this. First, the successive increases in salience conveyed by the peripheral uncertainty display ensured that changes in urgency are obvious and easy to perceive. Thus, participants could rely on noticing the changes and were not required to regularly check the uncertainty display. Second, the positioning of the display and the differences in representational fidelity may have contributed to the higher score. As such, the location of the light strip above the tablet computer allowed participants to sense the information peripherally, thereby reducing the information access effort (see Figure 2.7). Further, the increased abstraction afforded a faster cognitive processing of the information and required less cognitive effort (see Subsection 2.4.3). Despite the increase in abstraction which might have led to a reduction in comprehensibility and learnability (MacEachren, Roth et al. 2012), overall usability scores were higher as well. This is likely the consequence of extensive testing and the data-driven selection of each stimulus that preceded the user-based evaluation (see Section 6.1).

With emotional responses being a direct consequence of the perception of both instrumental and non-instrumental qualities (Thüring and Mahlke 2007), the recorded differences in positive and negative emotions comply with existing research and expectations. It can be argued that a display conveying the uncertainties of a system is unlikely to evoke positive emotions. Indeed, the scores for the emotional subscale regarding both display types are low compared with the overall appraisal. This coincides with equally low scores regarding the non-instrumental subscale, particularly for commitment. Thus, the display design may be improved through a more emotional and futuristic but less technical design (Thüring and Mahlke 2007). The lower values for commitment are in line with previous research (Minge and Thuering 2016). Thereby, the short usage duration and artificial setup may have been insufficient to form a bond between user and display (Minge and Thuering 2016).

The scores for Cronbach's alpha indicate the internal consistency of the PADE questionnaire, implying its overall reliability. This complies with expectations, as most items were adopted from questionnaires that had previously been tested for internal consistency (Minge and Thuering 2016). The results do, however, highlight that the subscale comprehensibility falls below the critical value of 0.70 (based on responses regarding the digital display). As Cronbach's alpha is a function of the number of questionnaire items (Trobia 2008), the value tends to increase by adding items. With a total of three items, the subscale contains relatively few items. Further, Cronbach's alpha only dropped below 0.70 for one of the surveys. Finally, the item adds the aspect of glanceability derived from Shami, Lshed and Klein (2005) to the questionnaire, which can be considered an integral part of peripheral awareness displays (see Chapter 5). It can therefore be argued that the item should remain in the questionnaire unless future evaluations also indicate insufficient values for Cronbach's alpha.

6.2.3.4 Limitations

The discussed results of the user-based evaluation are confined by the following limitations. Responses to several items of the PADE questionnaire likely change with extended usage. As discussed, particularly items that require the formation of a bond between user and product, such as those belonging to the scale commitment, may receive higher scores. The establishment of a bond is further impeded through the artificial setup. Thus, future research should focus on evaluating the uncertainty display after extended usage in real-life situations. Further, the experimental scenario did not include errors and therefore questions regarding the scale error visibility should be evaluated in future research efforts.

Additionally, the chosen implementation for the digital display likely affected the results. Instead of using numerical indications which require higher cognitive effort to link the information to reality (see Figure 2.10), the digital display could have employed abstract signifiers similar to the peripheral uncertainty display. However, the increase in abstraction – and change in position – are the main features that distinguish the peripheral from the digital display. These very features resulted from the context of use analysis (see Section 6.1) and were implemented to tailor the uncertainty display to the usage context. The results of the user-based evaluation highlight that these context-specific adaptations led to improvements in user experience. The abstract signs were similarly well understood as the numerical indications and improved the peripherality of the display (see Table 6.12). Equipping the digital display with the same features would likely improve its UX score.

6.3 Summary and Conclusion

This chapter successfully applied and evaluated the PADE framework and process developed in Chapter 5. Aimed at removing the derogative effects of conveying uncertainties using a digital display (see Chapter 4), a novel approach to uncertainty communication using a peripheral awareness display was presented.

The systematic development guided by the process depicted in Figure 5.8 ensured the thorough consideration of the usage context. As such, the uncertainty display was not regarded as a stand-alone instrument, but rather acknowledged as one of many instruments in the complex environment of a vehicle cockpit. Thus, user tasks were analysed and likely unoccupied sensory channels were selected to convey the uncertainty information. This is in contrast to existing research (Beller, Heesen and Vollrath 2013; Helldin et al. 2013) that presented the uncertainty information using a digital display located in the instrument cluster, thereby addressing the (focal) visual channel which is likely preoccupied with NDRTs (see Subsection 6.1.1), particularly with increasing driving automation (see Subsection 2.5.2). The improved UX score relative to a digital display highlights the positive implications of a user-centred, context-dependent design.

Further, the developed process (see Figure 5.8) allowed for resourceful testing of designs. While the design was not yet technically implemented and thus changes could be realised with minimal consequences for costs and effort, the heuristic evaluation afforded the improvement of several conceptual aspects based on a low-fidelity description of the interface design. Thus, the more elaborate user-based evaluation could employ a more refined design.

The UX score of the developed display indicates that its operating principles are easy to learn and understand, the conveyed information is easy to perceive and changes in urgency are noticeable, with appropriate levels of interruption. In contrast, the appeal of the display has room for improvement.

As particularly the instrumental qualities of the display are of interest when evaluating its implications for driving safety, the most recent iteration of the display (see Figure 6.13) can serve as a suitable implementation. Whereas the current chapter focused on developing and evaluating a display concept that affords the peripheral awareness of uncertainty communication, the subsequent chapter builds on this work and assesses the implications of peripheral uncertainty communication for human-automation interaction.

CHAPTER 7

Implications of Peripheral Uncertainty Communication

Abstract Besides promising advances in safety and fuel economy, driving automation systems afford users flexibility to engage in NDRTs. This entails a change in task hierarchy, whereby the significance of the driving task decreases and other, now secondary tasks, move to the centre of users' attention. Abandoning the driving task, even if only temporarily, inadvertently compromises safety if the system design requires a fallback-ready user (SAE level 3). Specifically, neglecting the driving environment leads to a reduction in SA and impedes successful takeovers, which is further intensified through inappropriate trust (see Chapter 2). Conveying system uncertainties was shown to support trust calibration and promote safer takeovers (see Chapter 4). However, using a digital display positioned in the instrument cluster requires users to regularly inspect the display for changes, leading to increased workload and promoting the likelihood of missed events. Counteracting the negative effects of such displays, a peripheral awareness display consisting of a light strip and a vibro-tactile seat matrix was developed to allow the perception of uncertainty information without visual focus (see Chapters 5 and 6). This chapter presents a driving simulator experiment that investigates the implications of the developed display on takeover performance, trust calibration, gaze behaviour, workload, NDRT performance, and physiological measures relative to the previously employed digital display. The peripheral awareness display allowed users flexibility to engage in NDRTs while decreasing the glance time off-road prior to critical situations, leading to improved takeover performance and a reduction in workload. Aspects of this chapter, in part verbatim, are published in Kunze, Summerskill et al. (2019b).

Background	Objectives	Method	Results	Implications
Using a digital instrument cluster to convey uncertainties requires operators to regularly change their gaze to perceive changes, leading to increased workload.	Investigate the human factors implications of using a peripheral awareness display instead of a digital display for conveying uncertainties in the intended context.	Within-subjects driving simulator experiment involving 24 participants that experienced both the peripheral awareness display and the digital display.	Off-road glance time is reduced in critical situations, leading to improved takeover performance using the peripheral awareness display. Further, workload is reduced.	Using peripheral awareness displays to convey uncertainties affords users the flexibility to engage in NDRTs without compromising on driving safety.

Figure 7.1: Overview of Chapter 7

7.1 Method

A driving simulator experiment was conducted to investigate the implications of using a peripheral awareness display relative to a digital display for uncertainty communication. Specifically, the following research questions were addressed:

1. How does the display type (peripheral awareness display; digital display) affect operator trust and monitoring behaviour?
2. How does the display type affect takeover performance and quality?
3. How does the display type affect operator workload?
4. How does the display type affect physiological measures?

The experiment builds on work outlined in Chapters 4 to 6. As such, it employs the interaction concept depicted in Figure 6.13 as an implementation for peripheral awareness displays and the numeric heartbeat display outlined in Chapter 4 as an instance for digital displays. As previously investigated (see Subsection 6.2.3), both display implementations are characterised by good usability levels. Further, the selected digital display was shown to be useful in a dynamic driving scenario, as indicated by improvements to SA and trust calibration relative to a control group that had no such display available (see Chapter 4). Consequently, both displays are valid implementations for each display type that ensure the comprehensibility of the conveyed uncertainty information.

7.1.1 Participants

The experiment was completed by the same sample of participants as the study outlined in Subsection 6.2.3. As such, a total of 24 participants (14 female) with a mean age of 27.417 years (SD = 6.795, range: $21 - 55$) volunteered to take part in the experiment. Participants held a driving licence for an average time of 9.438 years (SD = 6.456, range: $3 - 37$) and stated to drive an average of 5686.125 miles per year (SD = 4126.792). None of the participants reported to have been diagnosed with heart disease, respiratory conditions, or visual deficiencies such as colour blindness. There was no overlap of participants with the heuristic evaluation or the experiment outlined in Chapter 4. Ethical approval was granted by the Ethics Approvals Sub-Committee of Loughborough University.

7.1.2 Experimental Design

The experiment was conducted using a within-subjects design with the independent variable display type (peripheral awareness display; digital display). Four uncertainty levels were implemented for each display (see Table 6.11). To prevent learning effects, participants experienced two scenarios differing in the temporal sequence of uncertainty sections and the lateral takeover position of the ego and lead vehicle (see figures 7.2 and 7.3). Consequently, a total of four combinations were tested:

1. Peripheral awareness display and scenario A
2. Peripheral awareness display and scenario B
3. Digital display and scenario A
4. Digital display and scenario B

In accordance with a Latin Square plan for repeated measures (Winer 1971), all participants experienced each level of both independent variables by completing two of the combinations above (either 1 and 4 or 2 and 3). To prevent order effects, the sequence of the two combinations was fully balanced. Consequently, the same number of participants experienced the sequences 1-4, 4-1, 2-3, and 3-2.

To reproduce more realistic failure rates of real-world systems, the experimental scenario did not prompt users to take over the driving task multiple times (see Figure 7.2). Rather, it follows the experiment outlined in Chapter 4, which employed a single automation failure to invoke the strongest effects of insufficient vigilance and increased complacency (Parasuraman and Manzey 2010; Davies and Parasuraman 1982). Participants were notified of a system failure with a TOR that was conveyed with 7 s TTC remaining (Gold, Damböck et al. 2013) using combined visual (red flashing of the instrument cluster) and abstract auditory cues (Bazilinskyy et al. 2018; Politis, Brewster and Pollick 2015a). The TOR communication method was similar for all test combinations.

Whilst the automation was activated, participants were instructed to engage in a visually demanding NDRT to replicate a realistic driving scenario (see Subsection 6.1.1) and to account for the influence of NDRTs in a broader context (see Figure 2.14). Replicating the experiment outlined in Chapter 4, the visual search task prompted participants to identify if a certain target is present within a field of distractors (see Subsection 3.2.3.4) (Treisman and Gelade 1980).

7.1.3 Dependent Variables

Measurements on trust, NDRT performance, gaze behaviour, takeover performance, operator workload, heart rate, HRV, and respiratory effort were taken to investigate the research questions (see Section 7.1). Similar to the methods applied in Chapter 4, trust was measured using single-item ratings on a percentage scale to allow for the repeated, non-distracting measurement without interfering with participant behaviour or interrupting the experimental scenario. Figure 7.2 depicts at which points during the simulation trust measures were taken.

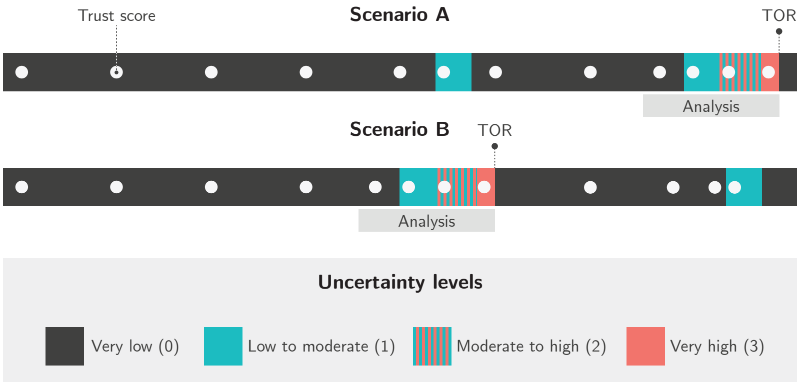


Figure 7.2: Timing of trust indications, TOR, and uncertainty cues for both experimental scenarios (Kunze, Summerskill et al. 2019b)

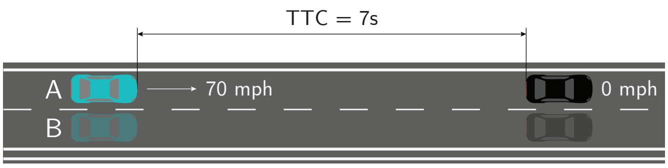


Figure 7.3: Driving situation when TOR was issued for scenario A (top) and B (bottom, less opaque) (Kunze, Summerskill et al. 2019b)

For each instance, the experimenter prompted participants to respond to the following question: 'To what percentage do you currently trust the automated system to safely perform the driving task?' Gaze behaviour was measured using eye tracking. In contrast to Chapter 4, the focus of this experiment was not to establish how frequently participants changed their gaze towards the FRD (corresponding to monitoring frequency, Equation 3.3), but rather how much time is spent on each AOI, particularly the uncertainty display, FRD, and NDRT. Consequently, the monitoring ratio was calculated using Equation 3.4. Besides, the NDRT performance served as an indirect indicator for monitoring behaviour. Takeover performance was measured using TTC, maximum lateral acceleration, and TTT. Measures on subjective operator workload were taken using the NASA-TLX workload scales.

Supplementing the dependent variables taken in Chapter 4, HRV and respiratory rate were measured as indicators for stress. Using Equation 3.6 and ECG data, RMSSD of R-R intervals were calculated as measures for HRV (see Subsection 3.2.3.7).

7.1.4 Apparatus

The experiment was conducted in a soundproof laboratory using the driving simulator described in Subsection 3.2.3.1. Following the implementation for a peripheral awareness display discussed in Chapter 6, the setup was complemented by a light strip and a vibro-tactile seat matrix (see subsections 6.1.4 and 6.1.5, figures 6.6 and 6.12). The instrument cluster was used to show the digital display consisting of a heartbeat animation and a numeric indication for beats per minute (see Figure 4.2). The visual search task was performed on an Apple iPad Air 2 tablet attached to the centre console. Tobii Pro Glasses 2 were used to record eye tracking data at a sampling frequency of 100 Hz. The recorded eye tracking data was subsequently processed with the proprietary software Tobii Pro Lab (Tobii AB 2018b).

Physiological measures were taken and processed using a BIOPAC MP35 four channel data acquisition system and the corresponding software *AcqKnowledge* (BIOPAC Systems Inc. 2017). Sampling at 200 Hz, three analogue input channels were addressed with (a) a BIOPAC electrode lead set SS2L (BIOPAC Systems Inc. 2013), (b) a BIOPAC pulse photoplethysmogram transducer SS4LA (BIOPAC Systems Inc. 2018a), and (c) a BIOPAC respiratory effort transducer SS5LB (BIOPAC Systems Inc. 2018b). The electrode leads were connected with three disposable electrodes, whereby *GRND* was attached to the right ankle, *VIN+* was positioned on the left ankle, and *VIN-* was secured on the right wrist of the subject. The optical pulse sensor was wrapped around the participant's right index finger using a provided Velcro® strap. The respiratory effort transducer was positioned approximately 5 cm below the armpits of each subject and fastened with a nylon belt.

7.1.5 Procedure and Scenario

Upon arrival, participants were briefed about the experimental procedure and were asked to fill in a consent form as well as a demographic survey. Next, the experimenter provided a more detailed description of the automated driving system. As such, the fallibility of the system was highlighted and consequences of system failures, i. e. takeovers, were explained. Further, the TOR feedback was presented to participants and they were guided through the visual search task with a series of sample questions. All instructions were provided in a consistent, scripted manner to prevent effects of differences in wording on dependent variables (Körber and Bengler 2013).

Following the introduction, participants were asked to take place in the driving simulator, adjust their seat position, and commence with a practice scenario lasting approximately 15 min. After the practice scenario, participants were equipped with eye tracking glasses and transducers for physiological measurements. Each measurement channel was calibrated for 60 sec to establish internal hardware parameters such as gain, offset, and scaling.

Subsequent to the calibration process, participants received a detailed introduction about the first uncertainty display they would be interacting with and had a final opportunity to ask any question they might have. Then, the 20 min experimental scenario started with a conditionally automated driving system (SAE level 3) engaged. Simultaneously with the start of the simulation, the recording for each physiological variable commenced. In the experimental scenario, the automated system was immediately engaged and performed several lane changes and passing manoeuvres on a two-lane UK motorway with moderate traffic while maintaining a speed of 70 mph. The uncertainty levels varied linearly with the visibility range due to fog from 91.44 m (thick fog, very high uncertainty) to 1005.84 m (no/low fog, very low uncertainty). The changes in uncertainty regarding each experimental scenario are depicted in Figure 7.2. Figure 7.3 illustrates the takeover situation, which was modelled after recent automation failures in which a stopped vehicle caused crashes (Shepardson 2018; Stewart 2018). To avoid a collision, participants had to perceive the stopped vehicle ahead and perform an evasive manoeuvre. Ego and lead vehicle were positioned in the left lane for scenario A and in the right lane for scenario B. Albeit mirrored, the relative positioning of each vehicle was similar in both scenarios. To prevent learning effects, the vehicle types were randomised.

Immediately after the completion of each experimental scenario, participants filled in a NASA-TLX workload questionnaire (Hart and Staveland 1988). Following the completion of both scenarios, participants were asked to state their preference concerning the displays and were given the opportunity to elaborate on their experience within the scope of semi-structured interviews.

7.1.6 Data Analysis

Multiple measurements on dependent variables were taken for each participant, leading to non-independent data points. Following the discussion in Subsection 3.2.5, LMMs were performed using R and the package lme4 (Bates et al. 2018). Unless explicitly stated otherwise, random effects were implemented for subjects using intercepts. To account for the ordinal characteristics of the independent variable uncertainty level, it was treated as continuous. The display type as well as the scenario were, in contrast, added to the model as categorical variables. Residual plots were visually inspected to validate the assumption of homoscedasticity (see Subsection 3.2.5). The R package lmerTest (Kuznetsova, Brockhoff and Christensen 2018) was applied to approximate degrees of freedom and p values for t statistics based on Satterthwaite's method.

7.2 Results

The recorded data were processed using R and the results are subsequently presented in the order of the research questions they address (see Section 7.1).

Table 7.1: Results of the performed linear mixed effects models with approximated degrees of freedom and p values based on Satterthwaite's method

Variables	***Estimate***	***SE***	***df***	***t***	***p***
Subjective trust					
Uncertainty level	−19.175	0.780	548.000	−23.976	<.001***
Display type	−2.331	1.265	548.000	−1.843	.066
Scenario	−0.569	1.079	548.000	−0.528	.598
Uncertainty level : display type	1.924	1.131	548.000	1.701	.090
Solving rate of NDRTs					
Uncertainty level	−6.236	0.546	164.000	−11.423	<.001***
Display type	0.781	1.444	164.000	0.541	.589
Scenario	1.446	0.863	164.000	1.675	.096
Uncertainty level : display type	−0.853	0.772	164.000	−1.105	.271
Solving rate of NDRTs					
Subjective trust	0.252	0.020	176.208	12.771	<.001***
Display type	−0.366	1.018	164.693	−0.359	.720
Scenario	1.282	1.018	164.694	1.259	.210
Heart rate					
Uncertainty level	−0.655	2.477	68.000	−0.264	.792
Display type	2.193	2.477	68.000	0.885	.379
Scenario	1.150	1.752	68.000	0.656	.514
Uncertainty level : display type	−2.498	3.504	68.000	−0.713	.478
RMSSD					
Uncertainty level	−15.912	6.473	52.513	−2.458	.017*
Display type	−4.845	6.585	53.076	−0.736	.465
Scenario	−0.641	4.687	53.277	−0.137	.892
Uncertainty level : display type	7.511	9.223	52.189	0.814	.419
Respiratory rate					
Uncertainty level	−0.192	0.875	68.000	−0.220	.827
Display type	−0.934	0.8753	68.000	−1.067	.290
Scenario	0.868	0.619	68.000	1.403	.165
Uncertainty level : display type	0.883	1.238	68.000	0.714	.478

$^{*}p < .050$ $^{**}p < .005$ $^{***}p < .001$

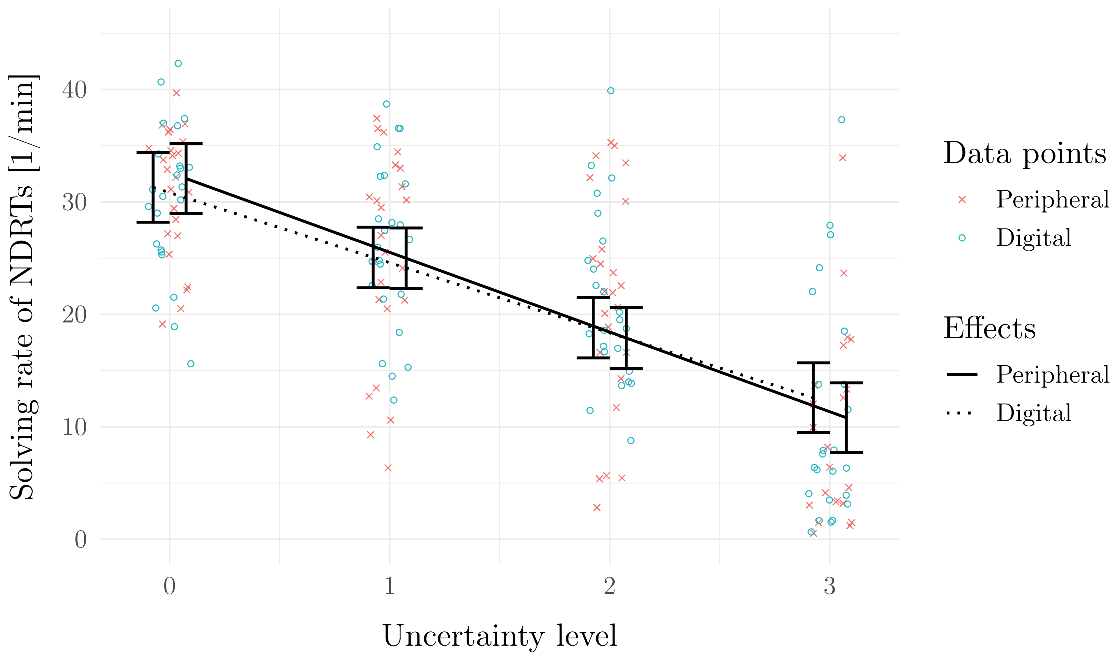


Figure 7.4: Effects of the performed linear mixed effects model predicting the impact of uncertainty level and display type on solving rate of NDRTs

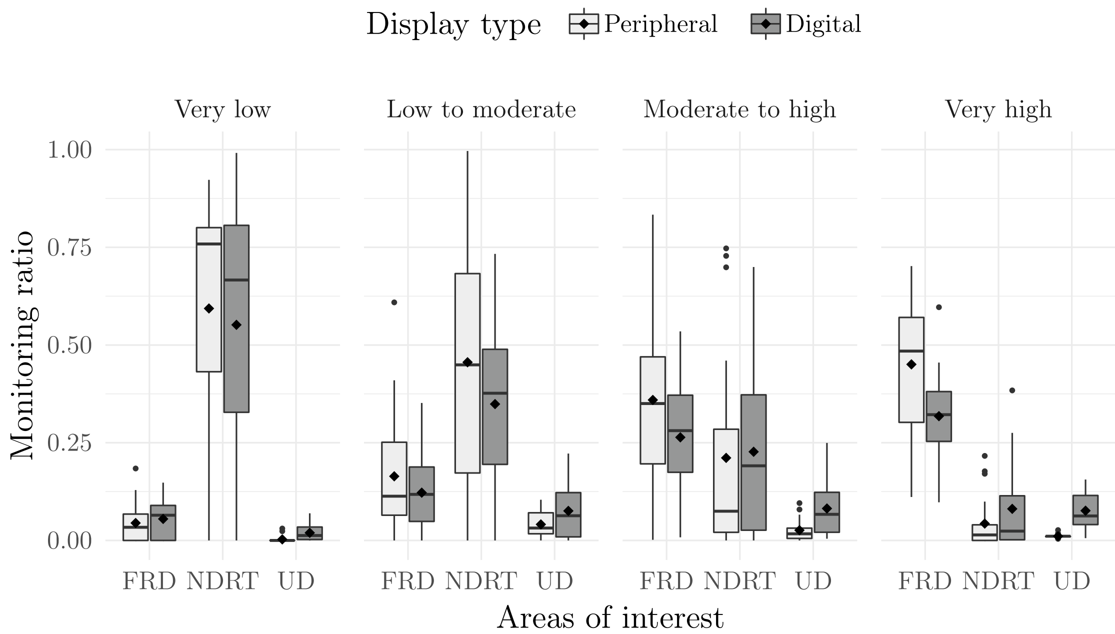


Figure 7.5: Monitoring ratio for each AOI depending on uncertainty level and display type (diamonds indicating mean values, bullets show outliers, UD: uncertainty display)

7.2.1 Impact on Trust and Monitoring Behaviour

Table 7.1 summarises the results of the performed LMMs. To assess the effects on the dependent variable subjective trust, the independent variables uncertainty level, display type, and scenario as well as the interaction of uncertainty level and display type were implemented as fixed effects. There was a significant main effect of uncertainty level on trust scores, $t(548.000) = -23.976, p < .001$. Subjective trust decreased with higher uncertainty, whereby gradient and intercept were not significantly different between display type.

The monitoring ratio regarding each AOI and the solving rate of NDRTs served as indicators for the monitoring behaviour of participants. Table B.7 summarises the descriptive statistics of the NDRT solving rate for each uncertainty level and display type. An LMM with the fixed effects uncertainty level, display type, scenario, and the interaction between uncertainty level and display type was performed on the dependent variable solving rate of NDRTs. There was a main effect of uncertainty level, $t(164.000) = -11.423, p < .001$. The display-specific trends visualised in Figure 7.4 suggest that the solving rate of NDRTs decreased with higher uncertainty for both display types. As both subjective trust and solving rate of NDRTs linearly varied with the uncertainty level, an additional LMM was performed to inspect the relationship between the two. The results indicate a significant main effect of subjective trust on the solving rate of NDRTs, $t(176.208) = 12.771, p < .001$, whereby the solving rate of NDRTs increased with subjective trust. Table B.5 and Figure 7.5 summarise the monitoring ratio regarding each AOI and uncertainty level for both display types. The recording device failed on two accounts, prompting the exclusion of four participants to retain a balanced design. Wilcoxon signed-rank tests were performed to assess the differences in monitoring ratio between groups. Regarding the highest uncertainty level, the monitoring ratio for the AOI uncertainty display was higher when using the digital display ($M = 0.076$, $SD = 0.046$) as compared with the peripheral awareness display ($M = 0.011$, $SD = 0.005$). At $p < .001$, it can be concluded that the factor display type significantly affects the monitoring ratio in the described context. In contrast, the monitoring ratio concerning the AOI FRD was higher for the peripheral awareness display during the same uncertainty level. Specifically, participants allocated less attention on the FRD when interacting with the digital display ($M = 0.318, SD = 0.119$) than with the peripheral awareness display ($M = 0.451$, $SD = 0.165$). At $p = .002$, it can be concluded that this difference is statistically significant.

7.2.2 Takeover Performance

Paired t tests were performed to assess differences in MTTC, maximum lateral acceleration, and TTT (see Subsection 3.2.3.1).

Table 7.2: Descriptive statistics for NASA-TLX workload subscales (Kunze, Summerskill et al. 2019b)

Workload subscales	Statistic	Peripheral	Digital
Mental demand	M	3.083***	5.250***
	SD	1.283	1.539
Physical demand	M	2.250**	3.083**
	SD	1.391	1.792
Temporal demand	M	2.917	3.792
	SD	1.442	1.615
Performance	M	5.750	5.167
	SD	1.225	1.435
Effort	M	3.125***	5.167***
	SD	1.513	1.659
Frustration	M	2.500*	3.583*
	SD	1.560	1.840

$^{*}p < .050$ $\quad$ $^{**}p < .005$ $\quad$ $^{***}p < .001$

When interacting with the peripheral awareness display, MTTC values were significantly higher ($M = 3.502\,\mathrm{s}, SD = 1.756\,\mathrm{s}$) compared with the digital display ($M = 2.496\,\mathrm{s}, SD = 1.572\,\mathrm{s}$), $t(23) = 2.509, p = .020, d = 0.604$. At $p = .591$, it can be concluded that the maximum lateral acceleration values did not significantly differ between the peripheral awareness display ($M = 1.657\,\mathrm{m\,s^{-2}}$, $SD = 0.818\,\mathrm{m\,s^{-2}}$) and the digital display ($M = 1.540\,\mathrm{m\,s^{-2}}$, $SD = 0.490\,\mathrm{m\,s^{-2}}$). TTT values indicate faster response times when using the peripheral awareness display ($M = 1.246\,\mathrm{s}$, $SD = 1.352\,\mathrm{s}$) as compared with the digital display ($M = 1.991\,\mathrm{s}$, $SD = 1.706\,\mathrm{s}$), however without statistical significance, $t(23) = 1.794, p = .086$.

7.2.3 Workload

The mean differences of individual workload subscales between display types were assessed using paired t tests. The results are summarised in Table 7.2. The mean scores for the digital display were significantly higher regarding the subscales mental demand ($t(23) = 6.397, p < .001, d = 1.529$), physical demand ($t(23) = 3.294, p = .00318, d = 0.520$), effort ($t(23) = 4.513, p < .001, d = 1.286$), and frustration ($t(23) = 2.306, p = .0305, d = 0.635$). In terms of performance, the peripheral awareness display received significantly higher scores, $t(23) = 4.513, p = .032, d = 0.437$.

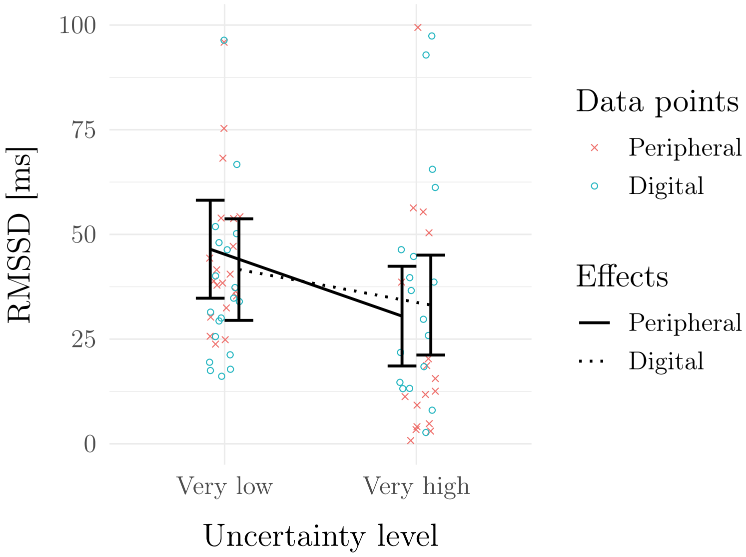


Figure 7.6: Effects of the performed linear mixed effects model predicting the impact of uncertainty level and display type on RMSSD

Table 7.3: Descriptive statistics of physiological measures taken during periods of very low and very high uncertainty for both display types

Uncertainty level	**Display type**	**Heart rate** [bpm]		**HRV** [ms]		**Resp. rate** [bpm]	
		M	*SD*	*M*	*SD*	*M*	*SD*
Very low	Peripheral	71.639	8.732	47.440	23.356	16.072	2.746
	Digital	73.832	10.849	38.528	18.911	15.138	2.831
Very high	Peripheral	70.984	12.778	31.802	34.616	15.879	3.912
	Digital	70.679	17.473	32.237	28.292	15.829	4.286

7.2.4 Physiological Measures

Table 7.3 summarises the descriptive statistics for heart rate, HRV, and respiratory rate. The analysis was limited to sections with extreme uncertainty levels (see Figure 7.2) to invoke the strongest effects. An LMM with the fixed effects display type, uncertainty level (very low; very high), scenario, and the interaction between display type and uncertainty level was performed on RMSSD measures. Outliers (RMSSD $\geq$ 113.016 ms) were removed as they substantially exceeded normal values which are considered to be around $M \pm SD =$ 27.3±22.2 ms for healthy adults and similar measurement durations (O'Neal et al. 2016). The observed outliers are likely the result of measurement errors as a consequence of temporarily unconnected electrodes. The results summarised in Table 7.1 indicate a significant main effect of the uncertainty level, $t(52.513) = -2.458, p = .017$. There was no main effect of display type or scenario and no significant interaction. For very low uncertainty levels, the average RMSSD when using the peripheral awareness display was higher ($M = 47.440$ ms, $SD = 23.356$ ms) compared with the digital display ($M = 38.528$ ms, $SD = 18.911$ ms). This was reversed for very high uncertainty levels, for which the digital uncertainty display was characterised by higher values ($M = 32.237$ ms, $SD = 28.292$ ms) than the peripheral awareness display ($M = 31.802$ ms, $SD = 34.616$ ms). Confirming the main effect of uncertainty level, the RMSSD values decreased with higher uncertainty levels for both display types. Figure 7.6 depicts the effects of the LMM to inspect the observed results. The gradients indicate that the drop in RMSSD was comparable for both displays, confirming the non-significant interaction and the main effect of *uncertainty level*. The recorded values for heart rate and respiratory rate indicate no significant effects (see Tables 7.1 and 7.3).

7.2.5 Interviews

Participant responses to semi-structured interviews were recorded and subsequently analysed in terms of common themes. Out of 24 participants, 20 preferred to use the peripheral awareness display (83.3 %), three favoured the digital display (12.5 %), and one was undecided (4.2 %). Participant responses showed that the peripheral awareness display afforded a higher reliance on the noticeability of changes without the need for active monitoring. One participant stated the following.

> The fact that I knew if it went from less to more light I would definitely notice it. I wasn't confident that I would notice if a number on the display changed. So I had to look more often for changes to make sure I did not miss anything.

The higher noticeability of the peripheral awareness display was commonly attributed to the simple coding of information and the positioning of the light strip. Moreover, the aesthetics of the light strip were favoured. Preference for the digital display was mostly justified with the more traditional location of driving-related information in the instrument cluster. In this context, participants also stated that this led to more active monitoring of the system in general. The participant that was undecided preferred individual aspects of both displays. The heartbeat (digital display) was liked because of its calming effect whereas the peripheral awareness display was favoured in terms of usability.

7.3 Discussion

The outlined results suggest a number of important implications regarding the use of peripheral awareness displays for uncertainty communication in the context of automated driving.

Further, the analysed data highlight several overall effects of uncertainty communication on human-automation interaction.

Irrespective of the display type, the subjective trust indications varied with the conveyed uncertainty level. An impact of the environmental indicator for uncertainty, fog, on the variations in trust is unlikely as it did not show a significant effect on trust scores in prior experiments (see Chapter 4). Thus, the recorded data provides further evidence for the positive effect of uncertainty communication on trust calibration, which is in line with prior work and the integral model depicted in Figure 2.14 (Beller, Heesen and Vollrath 2013; Helldin et al. 2013). Based on the linear relation between NDRT solving rate and subjective trust, the solving rate of NDRTs can be confirmed as a proxy measure for subjective trust levels, rendering regular interruptions that prompt users to indicate their trust in automation unnecessary.

Aside from the outlined conclusions that apply to uncertainty displays in general, the results imply a series of different implications depending on the display type. The recorded eye tracking data shows that participants directed a higher proportion of their gaze towards the FRD for very high uncertainty levels when using the peripheral awareness display. Similarly, the visual focus on the AOI uncertainty display was reduced for very high uncertainty levels when using the peripheral awareness display. This suggests that the peripheral awareness display afforded participants to perceive the uncertainty information with less visual focus, enabling them to direct more attention on the FRD prior to critical situations and thus supporting the acquisition of SA. As the off-road glance time contributes to an increase in crash risk (Seppelt, Seaman et al. 2017) and SA is critical for takeovers (Endsley 2017), the described differences were expected to result in safer takeovers. Indeed, the recorded MTTC values confirm that the peripheral awareness display supported drivers in taking over the driving task and performing an evasive manoeuvre in a safer manner.

Moreover, the NDRT task performance and eye tracking data indicate that both displays invoked appropriate reactions depending on the urgency level (see Subsection 5.2.4.2). In conjunction with workload data and physiological measures, the recorded eye tracking data allow the deduction of further conclusions that exceed the impact of the uncertainty display type on takeover safety. Focusing on the highest uncertainty level, the digital display kept participants more engaged in the NDRT (see Figure 7.5). This suggests that, at first, participants did not notice the change in uncertainty when using the digital display. In contrast, the multimodal feedback of the peripheral awareness display ensured an immediate perception of the change in uncertainty towards the uppermost level and prompted a reaction, i. e. to place more attention on the FRD. In reference to the reaction criterion (see Figure 5.5), the peripheral awareness display prompted a more accurate response (see Subsection 6.2.3). Despite feedback of a higher intensity, the peripheral awareness display did not increase operator stress to a higher degree than the digital display, as indicated by RMSSD values.

Independent from the uncertainty level, the peripheral awareness display afforded users to allocate more attention to AOIs other than the uncertainty display. This suggests that participants were better able to perceive the information without focal vision when using the peripheral awareness display, leading to a reduction in information access effort (see Figure 2.7).

NASA-TLX scores confirm the positive effect on workload in form of a reduction in perceived effort and physical demand. The scores further indicate a lower mental demand, reasons for which can be derived from the interview responses. Whereas the peripheral awareness display allowed users to constantly perceive the uncertainty information, users needed to remember the uncertainty level when using the digital display, likely increasing mental strain and providing a further explanation for higher monitoring ratios towards the AOI uncertainty display.

In addition to the presented conclusions that favour the use of a peripheral awareness display, the data also suggests room for improvement. Although the monitoring ratio regarding the AOI NDRT was higher for low to moderate uncertainty when using the peripheral awareness display, this did not translate to a significantly improved NDRT performance relative to the digital display. The same uncertainty level was also characterised by the highest monitoring ratio regarding the AOI uncertainty display for the peripheral awareness display. In this context, it is relevant to point out that, when the uncertainty changed to this level, the light strip was first switched on (see Figure 7.2 and Table 6.11). Thus, participants may have been briefly distracted by the light and were not able to fully engage in the NDRT. It should be investigated if this effect prevails when participants have familiarised themselves with the display. The fact that the monitoring ratio towards the AOI uncertainty decreased for higher uncertainty levels can be attributed to changes in stimuli. As humans are better able to perceive movement within their peripheral vision than, for instance, colour (see Subsection 2.4.3.2), the animation-based highest uncertainty level may have allowed users to more easily perceive the information without focal vision. The vibro-tactile feedback may have also rendered additional glances towards the uncertainty display unnecessary.

Additionally, the solving rate of NDRTs was not significantly higher for the lowest uncertainty level when using the peripheral awareness display. This might have been expected as a consequence of the reduced focus on the uncertainty display. Potentially, the monotonous visual search task resulted in low arousal, which, according to the Yerkes-Dodson law (Yerkes and Dodson 1908), may lead to overall decreased performance. As a consequence of the within-subjects design, participants might also have been more familiar with the automated driving system, leading to an increased non-observance of the digital uncertainty display (Large, Burnett, Morris and Muthumani 2017) and favouring higher NDRT solving rates. This increased the risk of missing changes when using the digital display, as indicated by longer reaction times following transitions to higher uncertainty levels (see Subsection 6.2.3.2).

The calculated RMSSD values as indicators for HRV and thereby stress also indicate detrimental effects of uncertainty communication. While the lower variation in R-R intervals indicates higher alertness and thus supports takeover performance, repeatedly high uncertainty levels might have negative effects on user experience (Yang et al. 2017). Once users are in a state of increased alertness, they should be given the opportunity to initiate the takeover themselves. This is still preferable over a sudden system failure in which case participants *must* take over the driving task.

Finally, discrepancies to the experiment outlined in Chapter 4 must be noted and discussed. Participants solved, on average, more tasks per minute than in the experiment outlined in Chapter 4. This may be a consequence of the within-subjects design utilised within this study. As each participant completed two scenarios, they were more experienced in completing the visual search task and thus, on average, performed better than in the previous experiment that used a between-subjects design. Further, workload scores regarding the digital uncertainty display varied between the experiments. This can be attributed to a different baseline. Whereas participants that interacted with the digital display in the experiment outlined in Chapter 4 had no point of reference due to the between-subjects design, participants of the current study compared the digital display with the peripheral awareness display, leading to differently calibrated scores.

7.4 Limitations

As previously mentioned, the detachment of electrodes, particularly on the wrist during high uncertainty levels, led to the exclusion of several data points regarding HRV. This was a consequence of the sudden change in hand position from the NDRT, i. e. the tablet computer mounted on the centre console, to the steering wheel as a preparation for an imminent takeover. Further, RMSSD values varied to a high degree between participants, as indicated by large standard deviations (see Table 7.3) and fairly heteroscedastic residual plots. Therefore, the reported HRV values should be used with caution.

Moreover, differences in absolute monitoring ratios relative to the experiment outlined in Chapter 4 despite the use of the same (digital) uncertainty display suggest limited external validity (see Tables B.4 and B.5). This can be attributed to alterations in the experimental scenario, for instance traffic density (Gold, Körber et al. 2016). Further, the repeated measures design may have allowed users to increasingly familiarise themselves with the automated system, which is confirmed by higher trust scores and NDRT solving rates (see Tables B.1, B.2, B.6 and B.7), leading to a decreased focus on the FRD for low uncertainty levels. Nonetheless, relative validity can be assumed as the general trends agree. As such, both experiments reported an increased focus on the FRD with higher uncertainty. Further, the duration of fixations on the UD increased from the lowest to the highest uncertainty level in both studies.

External validity is further limited as the study only varied the fog density as an indicator for different uncertainty levels. Other factors such as unclear lane markings, traffic density, or other weather conditions may also affect dependent variables such as trust of participants. Further, the selected sample cannot be considered representative for the overall population of drivers, prompting the need for more research in this context. Additionally, different lighting conditions may impact the visibility of the peripheral light strip. Therefore, the peripheral awareness display should be tested in real-life driving scenarios.

7.5 Summary and Conclusion

In the near future, automated driving systems will invoke a significant shift in travelling experience, whereby currently secondary, non-driving-related tasks move to the centre of users' attention. For as long as automated driving systems require a fallback-ready user, however, this change in task hierarchy inadvertently entails a trade-off with driving safety. The experiment outlined in this chapter has highlighted that peripheral awareness displays for uncertainty communication can support users in keeping the balance between driving safety and the engagement in NDRTs. Specifically, eye tracking data highlighted that the peripheral awareness display required minimal visual engagement for the perception and comprehension of uncertainty information. This can be attributed to the following key attributes. First, the use of variables with maximum abstraction ensured the rapid processing of information. The selection and implementation of the abstract signifiers for several levels of uncertainty was the result of a systematic, user-centred development process (see Chapters 5 and 6). Second, the position of the light strip on top of the centre console afforded users to peripherally perceive the information. Third, the multimodal approach combining peripheral visual and vibro-tactile feedback ensured the immediate perception of critical information. Fourth, increases in saliency for higher uncertainty levels reinforced the urgency of the conveyed information.

In reference to the integral model which was consolidated from literature review findings (see Figure 2.14), the designed peripheral awareness display guided top-down attention through changes in colour hue and light size, thereby counteracting effects of complacency and supporting trust calibration. Consequently, the visual focus of participants was increasingly changed towards the FRD for increasing levels of urgency and therefore uncertainty. The recorded eye tracking data, NDRT performance, and subjective trust indications support this. Importantly, the peripheral awareness display also guided the bottom-up attention allocation when uncertainty levels changed, implemented as transitions with attention-grabbing features such as pulses or vibro-tactile feedback. This counteracts the challenge that even optimal top-down attention allocation strategies do not guarantee that all changes are perceived, particularly with highly reliable automated driving systems that fail only on very few occasions and therefore result in less monitoring. Combined, the stimuli of the peripheral awareness display afforded users to direct more attention towards the FRD and thus acquire SA when needed, leading to safer takeovers.

Future research efforts should focus on optimising the display design. As such, it should be evaluated how the NDRT performance can be improved for lower levels of uncertainty. Further, the display feedback could be combined with sensors detecting the current state of users. For instance, users that already actively monitor the driving scenario should not be provided with feedback indicating a requirement for increased monitoring. Stationary eye tracking systems positioned in the vehicle cockpit could be employed to implement this.

CHAPTER 8

Implications of Function-Specific Uncertainty Communication

Abstract Communicating the inherent uncertainties of SAE level 3 automated driving systems has shown a series of benefits for drivers. Time-specific knowledge of system uncertainties improves trust calibration and supports the build-up of sufficient SA levels prior to critical situations, leading to safer takeovers (see Chapter 4). Whereas the use of a digital display for uncertainty communication led to workload increments due to additional monitoring requirements, peripheral awareness displays were shown to attenuate these negative effects (see Chapters 5 to 7). It has yet to be explored, however, how a variation in uncertainty information detail affects user behaviour. While uncertainties were treated as a system-wide value in previous research efforts, the implications of conveying function-specific uncertainties are a matter of further interest (see Subsection 2.9.3). Semi-structured interviews in conjunction with a driving simulator study involving 20 participants were conducted to assess the qualitative and quantitative implications of conveying uncertainties with a higher functional detail. Overall, the increase in functional detail showed no significant effect on takeover performance metrics. Interview responses suggest an even split between participants that prefer the higher functional detail and those who do not or are undecided. Results of a content analysis of interview responses highlight that workload should be considered when increasing the functional detail of uncertainty information. Combined, the results imply that users should be given the option to increase (or decrease) the functional detail of uncertainty information. Aspects of this chapter, in part verbatim, are published in Kunze, Summerskill et al. (2019c).

Background	Objectives	Method	Results	Implications
Increases in functional specificity allow users to gain more detailed knowledge about the uncertainties of each system function. This is unexplored in the context of driving.	Explore qualitative and quantitative implications of increasing the functional detail of uncertainty information in the context of automated driving.	Within-subjects driving simulator experiment involving 20 participants that interacted with a system-wide and a function-specific uncertainty display.	Overall, the functional detail did not affect takeover metrics. Preference scores were evenly split across displays, with workload being a critical requirement.	Ambiguous results suggest that users should be given the option of increasing (or decreasing) the functional detail of uncertainty information.

Figure 8.1: Overview of Chapter 8

8.1 Related Work

This chapter draws on the discussion surrounding trust contagion effects in complex systems, i. e. systems which have more than one automated aid (see Subsection 2.7.1.2). Two opposing theories regarding the behavioural implications of interacting with several automated aids differing in reliability exist, namely system-wide and component-specific trust theory. The former predicts that users ignore the reliability differences between the individual aids and merge their trust across the whole system. In contrast, the component-specific theory states that users can accurately distinguish the reliability levels of several aids and are able to adjust their behaviour accordingly. Results of the previously discussed experiments (see Subsection 2.7.1.2) suggest that the observed user behaviour does not confirm with either theory but rather lies on a continuum spanning between the two extremes.

For aids supporting decision selection (Meyer 2001, 2004; Chancey et al. 2017; Geels-Blair, Rice and Schwark 2013), a drop in the reliability of one aid was shown to affect the interaction with all other aids, presumably the result of trust contagion across the complex system. Nonetheless, the performance when interacting with fully reliable aids is significantly better compared to the performance when using unreliable aids. Thus, contagion effects are measurable but not sufficiently strong to completely overshadow perceived differences in reliability. Further, explicitly presenting information about the reliability (or uncertainty) of each aid alleviates the contagion effects (Rice and Geels 2010). Thus, conveying system reliability (or uncertainty) information counteracts trust contagion effects and may lead to more appropriate trust in automation (see Subsection 2.7.1.1).

Existing research on trust contagion effects and trust in complex systems has focused on aids supporting the decision selection of participants. In automated driving systems with SAE level 3 and higher (SAE International 2018), the complete DDT can be performed by an automated aid, thus covering all categories of automation, from information selection to action implementation (Parasuraman, Sheridan and Wickens 2000).

Previous studies have investigated the communication of the overall system uncertainties, hence the uncertainties regarding the complete DDT (Beller, Heesen and Vollrath 2013; Helldin et al. 2013). For this approach, a generic metric for the current uncertainty is calculated from the interplay of sensors, actuators, and software (see Subsection 2.5.4). An increase in uncertainty in one of several sensors or actuators is thereby indicated by an increase in the overall system uncertainty. Using this approach, the individual uncertainty figures are not available to the user.

An alternative approach could present the uncertainty information with a higher functional detail. As a consequence, users may have a better awareness of the automated system and can build their mental model accordingly (see Section 2.7). Thus, users may not only know *that* something is awry, but they are also pinpointed to a more specific reason for that uncertainty (Koo et al. 2015).

To the knowledge of the author, no existing publications investigated differences in functional detail of uncertainty information regarding automated driving systems (see Subsection 2.5.5). Further, the qualitative implications of conveying function- or component-specific uncertainty information have yet to be explored to answer questions regarding subjective preference and users' reasoning behind it.

8.2 Method

A driving simulator study followed up by a semi-structured interview was conducted to investigate the following research questions:

1. How does takeover performance differ when uncertainty is communicated in a function-specific manner compared with an overall system indication?
2. How do system failures regarding the lateral or longitudinal component of the driving task affect takeover performance under consideration of the functional detail of uncertainty information?
3. How does the communication of uncertainties with a higher functional detail compare with the presentation of the overall system uncertainties in terms of preference and what are the reasons for preferring one over the other?

As discussed in Subsection 2.9.3, the DDT can be split into a lateral and longitudinal component (SAE International 2018). This distinction is likely familiar to current users of vehicles equipped with ACC (for longitudinal control) and lane assist (for the lateral component). Other drivers who have not previously used advanced driver assistance systems are also likely to find the functional distinction intuitive as the controls for lateral and longitudinal control are spatially divided in all on-road vehicles. Hence, the provision of uncertainty information regarding a lateral as well as a longitudinal component will be used as an implementation for an increase in functional detail.

8.2.1 Participants

A total of 20 participants (9 female) with an age ranging from 23 to 48 years ($M = 28.350, SD = 5.761$) completed the experiment. Valid driving licences were held by all participants ($M = 9.050$ years, $SD = 6.004$). On average, participants reported to drive 4490.000 miles per year ($SD = 3403.543$). A total of six participants reported to have used vehicles with SAE level 1 or higher (SAE International 2018), whereas 14 participants reported to have no experience with driving automation functions. Four participants indicated to have used vehicles equipped with both lane assist and ACC. Two participants stated to have used either ACC or lane assist. Participants received no monetary compensation for taking part in the study. Ethical clearance was granted by the Ethics Committee of Loughborough University.

8.2.2 Design

The experiment was conducted using a within subjects design with the independent variables takeover scenario (curve section; straight section, see Figure 8.5) and functional detail (system-wide uncertainty; function-specific uncertainty with the two functions lateral and longitudinal vehicle motion control). The takeover scenario was varied in order to (1) prevent learning effects and (2) to investigate the impact of function-specific uncertainty communication on takeovers due to a failure of the (a) lateral and (b) longitudinal vehicle motion control. The experimental design resulted in four test combinations:

1. System-wide uncertainty display and takeover in a curve
2. System-wide uncertainty display and takeover on a straight section
3. Function-specific uncertainty display and takeover in a curve
4. Function-specific uncertainty display and takeover on a straight section

In accordance with a Latin Square plan for repeated measures (Winer 1971), all participants experienced each level of both independent variables by completing two of the combinations above (either 1 and 4 or 2 and 3). To prevent order effects, the sequence of the two combinations was fully balanced. Consequently, the same number of participants experienced the sequences 1-4, 4-1, 2-3, and 3-2.

For the takeover on a straight section, fog was selected as an environmental indicator for uncertainty as it is likely relatable for novice users of automated driving systems. Further, fog may reduce the reliability of lidar sensors as small water droplets scatter light pulses (Rasshofer, Spies and Spies 2011). Fading lane markings were used to indicate the environmental reason for uncertainty in the curve takeover scenario.

8.2.3 Dependent Variables

Interview responses were recorded and transcribed to allow the extraction of codes as part of a content analysis (see Subsection 3.2.4.2). Quantitative data was collected to augment the statements during the interviews. As such, measures relating to driving performance were recorded, including TTC, TTT, and lateral as well as longitudinal acceleration values.

8.2.4 Apparatus

For the duration of the experiment, participants were seated in a mock-up Land Rover cockpit (see Subsection 3.2.3.1). They faced three 160 cm × 90 cm projection screens that were inclined towards the driver to generate a 135° horizontal field of view. The software STISIM Drive 3 was used for the simulation. The simulated vehicle was equipped with automated transmission and showed the speed in miles per hour. In addition to the main simulation screens, a programmable instrument cluster with a resolution of 1280 × 480 pixels presented the uncertainty information to the user.

8.2.5 Uncertainty Visualisation

To further draw on the familiarity with existing vehicle functions such as ACC, the uncertainty of the system was displayed using symbols corresponding to standardised ISO icons and existing OEM solutions. Figure 8.2 depicts the ISO symbols for ACC, steering failures, and lane keeping assistance systems. Figure 8.3 shows a screenshot of the instrument cluster of a Tesla Model S.

While the indication for longitudinal vehicle motion control was found to be univocal (see Figure 8.2a), either an icon depicting a steering wheel (see Figure 8.2b) or one showing a vehicle within lanes (see Figure 8.2c) could be used for lateral vehicle motion control. To avoid confusion for situations in which a vehicle is in-between lanes, for instance during a lane change, the more generic icon of a steering wheel was used for lateral control. For the overall system uncertainty, a schematic vehicle in combination with a stylised computer circuitry was designed.

Figure 8.4 depicts the screens used for each experimental scenario. To indicate no uncertainty, the icon was coloured in a blue hue, similar to the colour scheme used in current vehicles equipped with automated driving systems (see Figure 8.3). Once the

(a) ISO 7000-2580 **(b)** ISO 7000-2441 **(c)** ISO 7000-3128

Figure 8.2: ISO symbols for ACC (a), steering failure (b), and lane keeping assistance system (c) (ISO 2017)

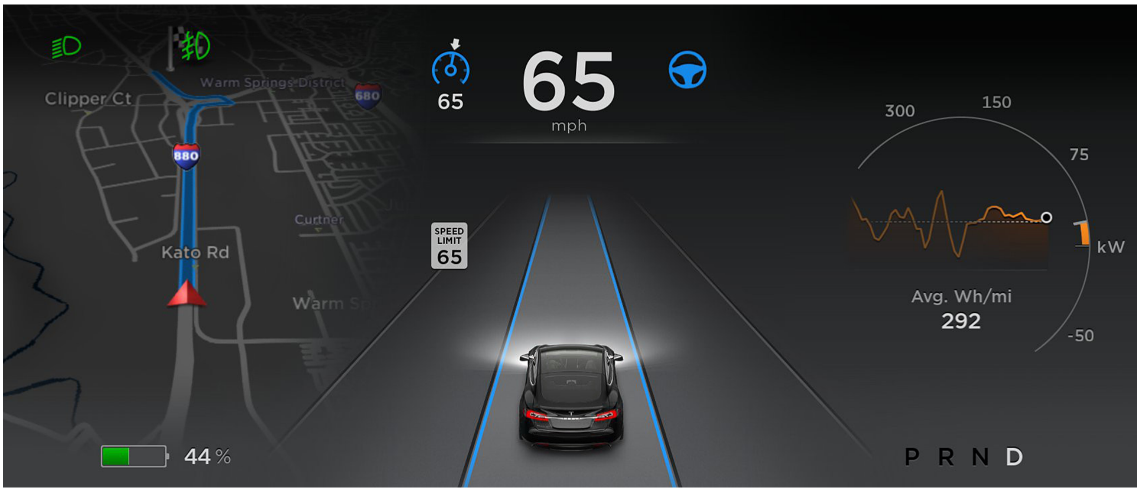


Figure 8.3: Instrument cluster of Tesla Model S vehicles (Consumer Reports 2015)

uncertainty increased, the colour of the respective icon was changed to a red hue and the icon was animated to show a shaking movement as a further metaphor for uncertainty.

8.2.6 Procedure and Scenario

Upon arrival in the driving simulation laboratory, participants received a briefing about the experiment including information about the type of collected data, the implications of the study, and an introduction to automated driving. After filling in a consent form and a demographic survey, the participants were seated in the driving simulator and completed a 15-minute practice scenario which involved multiple situations that required emergency braking and/or evasion manoeuvres. Following this, participants were exposed to two test conditions (see Subsection 8.2.2), using both functional detail levels (system-wide; function-specific) and experiencing both takeover scenarios (curve; straight section). Before each simulation run, participants were introduced to the display they were about to interact with. Each simulation scenario lasted approximately 10 minutes and involved an emergency takeover.

The simulation scenario implemented an SAE level 3 (SAE International 2018) automated driving system that was initially engaged, driving at a constant speed of 70 mph. Participants were instructed to monitor the automated driving behaviour and intervene if deemed necessary. For each scenario, one silent system failure prompted drivers to take over the driving task. Figure 8.5 depicts the driving situations at the time of the system failure, whereby each was implemented with 4 s TTC remaining. TTC was reduced compared with the previously outlined experiments (see Chapters 4 and 7) to account for the absence of competing tasks demanding visual attention.

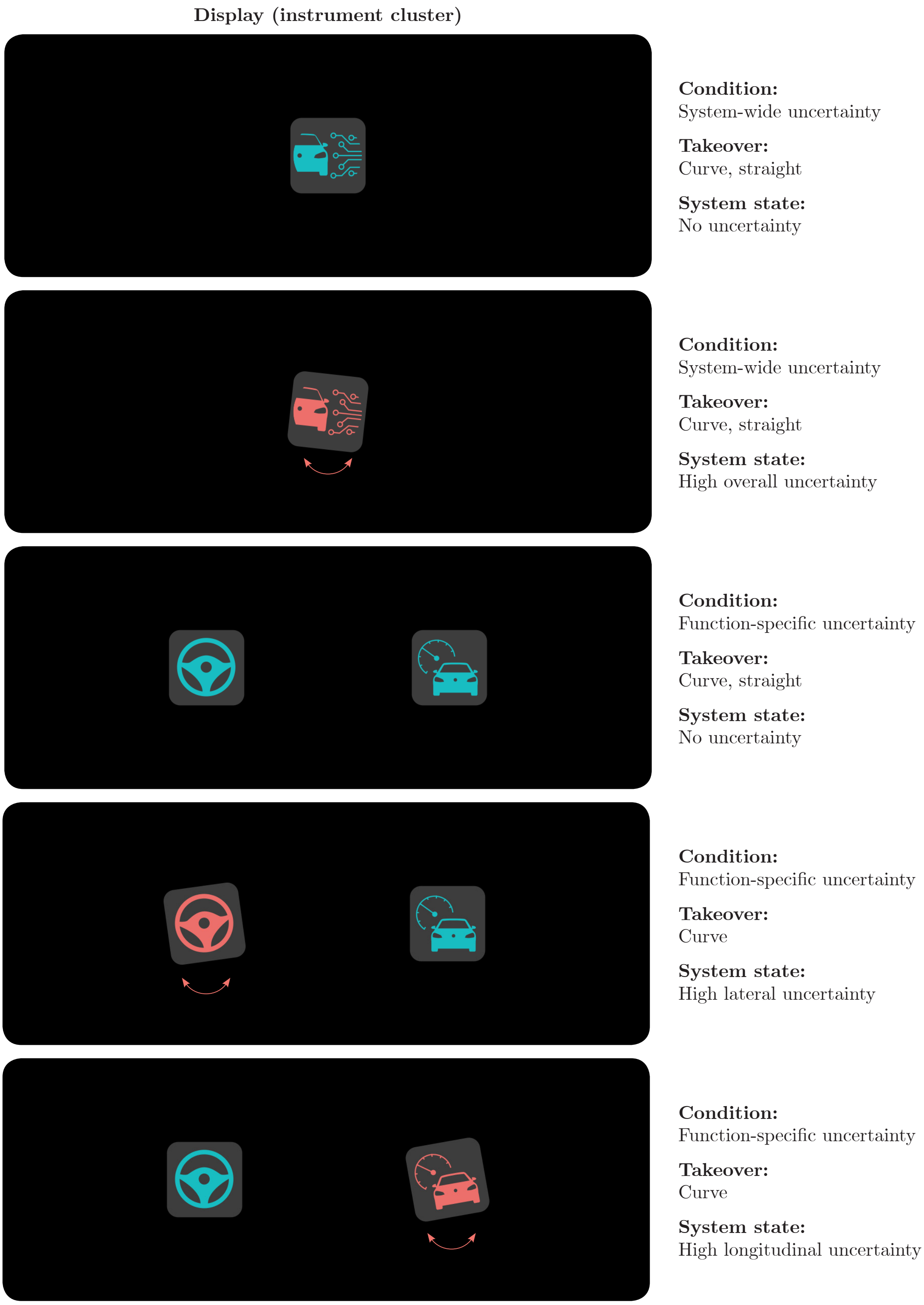
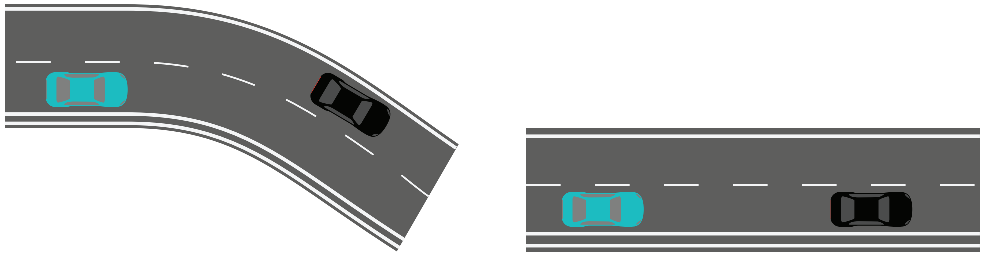


Figure 8.4: Screenshots of the instrument cluster for each experimental condition (arrows are included to illustrate icon movement) (Kunze, Summerskill et al. 2019c)



(a) Takeover in a curve **(b)** Takeover on a straight section

Figure 8.5: Driving situation at the time of system failure (4 s TTC to leading vehicle)

In one condition, the system failure occurred as the vehicle was entering a curve (see Figure 8.5a) with a curvature of 0.002 1/ft (easy right turn). In the second condition, the driving system failed as a leading vehicle was approached on the right lane. In order to prevent a crash, participants were required to take over the driving task and perform a braking and/or evasive manoeuvre. Contrasting other experiments presented in this thesis, the failure was not explicitly communicated to the driver. While automated systems are likely aware of their current sensor states and overall uncertainty, they might not detect the specific situations in which they actually perform incorrectly, referred to as *silent* failures (Louw, Kuo et al. 2019). Thus, the uncertainty of the system may not be sufficient to invoke a TOR but could still lead to erratic driving behaviour. Within this experiment, participants were required to notice that the vehicle was not braking to avoid a crash with the leading vehicle (straight section) or observe that the system was not steering the vehicle properly into a curve (curve takeover). Participants were instructed that a failure of either function would cause the complete automated system to shut down, i. e. drivers were required to takeover the complete DDT. Following the completion of both scenarios, the experiment was concluded with a semi-structured interview according to a prespecified interview guide (see Subsection 8.2.7).

8.2.7 Interview Guide

The study objectives specify the differences in subjective preference between the two uncertainty displays varying in functional detail as an overarching topic. The interview guide therefore included questions regarding both aspects of interest. Specific reasons for given responses, however, were expected to contain a wide spectrum of different answers, thereby impeding the use of survey questionnaires. Semi-structured interviews were selected to allow the researcher sufficient control about the interview direction without limiting the information detail provided by participants (see Subsection 3.2.4.1). Table 8.1 summarises the framework used for the interviews.

Table 8.1: Interview guide for gaining quantitative and qualitative information about the participants' interaction with uncertainty displays varying in functional detail

Question	**Rationale**
Tell me about how you experienced automated driving?	Introduction to establish rapport
Tell me about the display in the instrument cluster in the last part of the study	Evoke thought process about interaction with uncertainty display.
What differences did you notice compared with the display you used first?	Clarify differences between displays as a build-up for following questions.
Did your expectations regarding situations in which the system may fail vary between the two displays?	Closed question to verify that participants used the function-specific display as intended, i. e. participants should have expected failures to be me more likely in certain situations
(*If yes*) How did your expectations vary?	Insight to rationale for previous answer
Would you rather use the system with one or two indications for uncertainty?	Closed question to quantify preference
Tell me what you liked about display A/B.	Investigate reasons for preferring one over the other
Do you have any further remarks?	Debrief and give an opportunity to mention anything that has not been covered with previous responses

8.2.8 Data Analysis

Quantitative driving performance data including TTC, TTT, and lateral acceleration values were analysed using LMMs to account for the impact of the variable takeover scenario and the repeated measures design (see Subsection 3.2.5). Response counts of closed questions (see Table 8.1) were evaluated using chi-squared tests. Interview transcripts were processed by applying content analysis, whereby coding categories were extracted directly from the data (see Subsection 3.2.4.2). Interview responses were transcribed and subsequently thematically analysed following the procedure described in Subsection 3.2.4.2 (see Table 3.2 and Section A.6 for a sample transcript and analysis). Using the conventional approach (Hsieh and Shannon 2005) to content analysis, the interview transcript was increasingly abstracted. First, meaning units were extracted from each transcript and further condensed to allow the definition of codes. The deduced codes were subsequently grouped to form broader categories.

Table 8.2: Results of the performed linear mixed effect models with approximated degrees of freedom and p values based on Satterthwaite's method

Variables	*Estimate*	*SE*	*df*	*t*	*p*
MTTC					
Functional detail	0.290	0.406	35.000	0.714	.480
Takeover scenario	1.149	0.406	35.000	2.828	.008**
Driving experience	0.043	0.025	35.000	1.756	.088
Functional detail:TOR scenario	−0.360	0.579	35.000	−0.620	.539
TTT					
Functional detail	−0.437	0.486	30.237	−0.898	.376
Takeover scenario	0.643	0.486	30.237	1.324	.196
Driving experience	−0.317	0.034	17.000	−0.922	.370
Functional detail:TOR scenario	0.273	0.804	17.000	0.340	.738
Maximum lateral acceleration					
Functional detail	−0.577	0.735	32.976	−0.785	.438
Takeover scenario	−3.119	0.735	32.976	−4.242	<.001***
Driving experience	−0.040	0.049	17.000	−0.805	.432
Functional detail:TOR scenario	0.764	1.148	17.000	0.665	.515

*$p < .050$ **$p < .005$ ***$p < .001$

8.3 Results

Table 8.2 summarises the results of the performed LMMs. The factor driving experience was implemented as an additional fixed effect to account for its potential impact on the usefulness of increases in functional detail (Zeeb, Buchner and Schrauf 2015). Driving experience was input as the time in years since the applicable driving licence was issued. The impact of functional detail and takeover scenario on MTTC was assessed using an LMM under consideration of the interaction between the two variables. There was a main effect of the variable takeover scenario, $t(35) = 2.828, p = .008$, whereby the curve takeover scenario was characterised by lower MTTC values ($M = 1.295\,\text{s}, SD = 0.627\,\text{s}$) than the takeover on a straight section ($M = 2.264\,\text{s}, SD = 1.116\,\text{s}$). There was no main effect of the functional detail and no significant interaction between the independent variables. When taking over in a curve, participants using the uncertainty display with a higher functional detail achieved, on average, a MTTC of 1.394 s ($SD = 0.571\,\text{s}$). Participants that only had the system-wide uncertainty information available avoided a crash with 1.195 s ($SD = 0.694\,\text{s}$) TTC remaining. As predicted by the main effect of the variable takeover scenario, MTTC values for the takeover on a straight section were higher irrespective of the functional detail, with 2.253 s ($SD = 1.042$ s) for one uncertainty indication and 2.275 s ($SD = 1.243$ s) when uncertainty information was available for both lateral and longitudinal vehicle motion control.

An additional LMM was performed on TTT values using the same independent variables and interaction. The model returned no significant effects (see Table 8.2). Initial data suggested that when using the function-specific uncertainty display participants only took over the function for which a high uncertainty was indicated. Specifically, participants might only provide steering input when the uncertainty level for lateral vehicle motion control was high but failed to takeover the longitudinal vehicle motion control, i. e. braking or accelerating. Thus, the analysis also considered whether drivers only took partially over or if they took over both longitudinal and lateral control. A partial takeover was assumed if participants failed to provide input for the driving function that did not cause the system failure, i. e. if no brake or acceleration input was provided during the curve takeover and no steering input was recorded during the straight takeover. A time frame of 4 s following the system failure was chosen as this was the TTC at the time at which the automation malfunctioned. Across both takeover scenarios, 14 out of 20 participants provided steering and braking/acceleration input, irrespective of the display they interacted with. Six participants only provided input for the vehicle function that failed. For failures in a curve, 9/10 provided braking/acceleration input when using the system-wide uncertainty display, whereas the ratio was 7/10 for the function-specific display. Failures on a straight section prompted 5/10 participants to generate steering input when only having one indication available. With two indications, 7/10 participants altered the steering angle. The results of a chi-squared test with Yates' continuity correction indicate statistically non-significant differences, $\chi^2(1) = 0.146, p = .703$.

Further, an LMM was performed for the maximum lateral acceleration following the takeover (see Table 8.2). There was a main effect of takeover scenario, $t(32.976) = -4.242, p < .001$. The lateral acceleration in the curve takeover scenario was larger ($M = 4.141\,\mathrm{m\,s^{-2}}, SD = 1.367\,\mathrm{m\,s^{-2}}$) than that following the takeover on a straight section ($M = 1.404\,\mathrm{m\,s^{-2}}, SD = 1.777\,\mathrm{m\,s^{-2}}$). There was no main effect of functional detail and no significant interaction between the two independent variables. When interacting with the system-wide uncertainty display, the maximum lateral interaction averaged $4.388\,\mathrm{m\,s^{-2}}$ ($SD = 1.322\,\mathrm{m\,s^{-2}}$) in the curve takeover scenario and $1.352\,\mathrm{m\,s^{-2}}$ ($SD = 1.743\,\mathrm{m\,s^{-2}}$) in the straight takeover scenario. In contrast, the maximum lateral acceleration averaged $3.894\,\mathrm{m\,s^{-2}}$ ($SD = 1.436\,\mathrm{m\,s^{-2}}$) in the curve takeover and $1.456\,\mathrm{m\,s^{-2}}$ ($SD = 1.904\,\mathrm{m\,s^{-2}}$) in the straight takeover scenario when interacting with the function-specific uncertainty display.

Table 8.3: Counts for each combination of responses to closed questions (values in brackets indicate percental value relative to all responses)

		Expectations differed	
		Yes	No
Preference	Two	8 (50 %)	2 (10 %)
	Undecided	2 (10 %)	0
	One	3 (15 %)	5 (25 %)

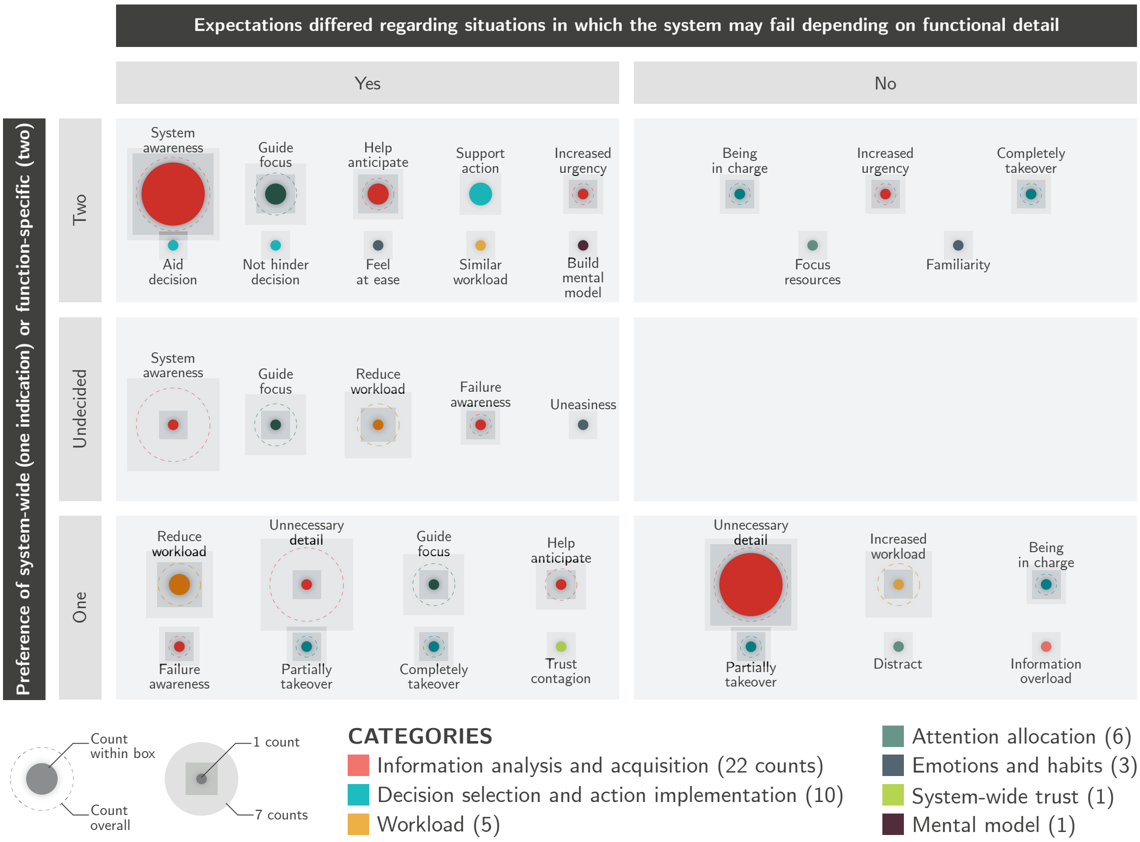


Figure 8.6: Codes extracted from interview transcripts, sorted into response combinations of the structured interview part (Kunze, Summerskill et al. 2019c)

Figure 8.6 summarises the content analysis results of the semi-structured interviews. The extracted codes were sorted according to the responses that were given to the closed questions, i. e. regarding the expectations of when the system might fail and subjective preference (see Table 8.1). The size of each bubble indicates how often the code was allocated to interview transcripts, whereby no single code was entered twice for any participant. The dotted line surrounding the bubble depicts the occurrence of the code across all participants, whereby the enclosed filled circle is sized relative to the number of mentions for the respective responses to the closed questions. System awareness, for instance, was mentioned seven times overall, but only once for participants which indicated that their expectations differed and who were undecided about which functional detail they prefer. The ratio between the size of the filled circle and that of the enclosing circle with a dashed border illustrate the relative importance of the code for each response combination of the closed questions. The colouring of each circle indicates the categories that the codes were grouped into. The categories *information analysis and acquisition* as well as *decision selection and action implementation* were derived from the levels of automation proposed by Parasuraman, Sheridan and Wickens (2000) (see Subsection 2.5.5). *Workload*, *attention allocation*, and *mental model* were extracted from the model presented in Figure 2.14. The category *system-wide trust* was established in reference to the discussion relating to trust in complex systems (see Subsection 2.7.1.2 and Section 8.1). Finally, the category *emotions and habits* was formed in reference to the user experience criteria presented in Section 2.3.

Table 8.3 summarises the counts for each combination of responses to the closed questions regarding expectations and preference (see Table 8.1). Overall, 10 participants, i. e. 50 %, preferred to use the function-specific uncertainty display with two indications. Eight participants would rather use the system-wide uncertainty display, whereas two participants were undecided. When expectations differed, participants preferred to use the function-specific uncertainty display (8-3). In contrast, the majority of participants who stated that their expectations did not differ preferred the system-wide display (5-2). The results of a chi-squared test indicate statistically non-significant differences, $\chi^2(2) = 4.725, p = .094$.

8.4 Discussion

The previously outlined results demonstrate several implications of function-specific uncertainty communication which lead to a number of recommendations for the design and development of uncertainty displays in a driving context. Non-significant effects of the uncertainty display's functional detail on any of the analysed driving performance metrics do, however, indicate that increasing the functional detail of information does not necessarily lead to improved takeover quality. Rather, MTTC, TTT, and acceleration values are on a similar level to that recorded for the system-wide uncertainty display.

While there was no significant effect of the functional detail, the effect of the variable takeover scenario was significant for MTTC and maximum lateral acceleration. This finding is in line with prior research (Sadeghian Borojeni et al. 2018). Lower TTC values and higher maximum lateral acceleration indicate that the takeover in a curve was more challenging for drivers. In contrast to the takeover on a straight section, the system failure at the start of a curve prompted drivers to provide a larger steering input in order to (a) control the vehicle through the curve and (b) avoid crashing into the leading vehicle. Consequently, the steering input led to a higher maximal lateral acceleration. Non-significant interactions between the two independent variables indicate that there was no impact of the functional detail dependent on the takeover scenario. Instead, the increase in functional detail had similar effects for both a failure at the start of a curve and a malfunction on a straight section. Potentially, the scenario was not sufficiently long for complacency effects (Parasuraman and Manzey 2010) and users were vigilant throughout the scenario, independent from the uncertainty information displayed on the screen (see Subsection 2.6.2.2).

Even though the collected driving data did not provide statistically significant differences between the system-wide and the function-specific uncertainty display, the results of the conducted content analysis regarding interview transcripts provide insights into why users may prefer one display over the other. The extracted codes demonstrate that subjective experience differed between participants despite the absence of a measurable effect on driving data.

With a total of 22 counts, statements falling into the category *information analysis and acquisition* were more frequent than those falling into any other single category. In particular, statements regarding *system awareness* (7 mentions) and *unnecessary detail* (7) were counted most often. The former was mostly used to argue in favour of having two uncertainty indications, whereas the latter was used to reason against it. Participants that favoured two indications were also more likely to have had different expectations regarding the situations in which the system may fail. The opposite was true for participants who preferred one indication: the expectations did mostly not differ between the displays (see Table 8.3). Thus, participants that reported the mentioned response combinations to the closed ended questions can be divided into those that took in the information and used it to improve their system awareness, and into those who ignored the fact that more detail was available and only used the red colour and wiggle animation of either icon as a cue for an increased takeover probability, irrespective of the failing function and situation. Driving performance data indicate that both strategies work equally well. Participants that preferred the display with two indications and reported a difference in expectations also used the information to guide their focus (2) and to anticipate (2) failures in certain situations, consequently leading to a support of the DDT (2). The fact that only two participants indicated that the function-specific display supported decision selection and action implementation may provide further rationale for why the function-specific display did not result in a superior takeover quality.

In contrast, participants who stated that their expectations differed but still preferred the uncertainty display with one indication correctly interpreted the different icons available in the function-specific display, but failed to use it to their advantage as a consequence of higher workload (2) and trust contagion (1) (see Figure 8.6). Further, the fear of accidentally only partially taking over the DDT when using the function-specific uncertainty display prompted users to prefer one indication. However, the analysis of steering and braking/acceleration data does not confirm that this was an issue (see Section 8.3). One out of 20 participants indicated that when using the function-specific display, the increased uncertainty of one driving function led to a distrust of the complete system, thus indicating trust contagion effects in line with the system-wide trust theory (see Subsection 2.7.1.2 and Section 8.1). While this does not provide evidence for the dismissal of the system-wide trust theory (Keller and Rice 2010), it illustrates that most users are at least not actively aware of merging their trust across functions.

Overall, there was no consensus about which display is preferred by users. Rather, preference was evenly split between the two displays. It is questionable if those who consider the information to be unnecessary would change their mind through an improved display design. However, those who reported that their expectations differed and were thus able to anticipate failures are potentially susceptible to adjust their preference through an improved design. In particular, display designs should aim at reducing operator workload which the monitoring of an additional system metric entails.

8.5 Limitations

The experiment presented in this chapter focused on the qualitative implications of increasing the functional detail of uncertainty information. Considering that basic themes can usually be extracted from six interviews and saturation occurs within approximately twelve interviews (Guest, Bunce and Johnson 2006), the sample size ($N = 20$) can be considered sufficient for the analysis of qualitative data. Although comparable research used the same or smaller sample sizes (Politis 2016), it may not have been sufficient for a rigorous quantitative analysis. Currently non-significant results might change with an increase in sample size. For instance, the effect of driving experience on takeover values should be addressed in future investigations. Further, the strong effect of the variable takeover scenario may diminish the effect of the functional detail, despite its consideration within the LMM. Thus, future studies should involve more participants without variation in the takeover scenario. As previously pointed out, the impact of the uncertainty display on takeovers may increase with longer scenario durations. As drivers were instructed to remain vigilant, complacency effects were unlikely to occur (Parasuraman and Manzey 2010). Uncertainty information does, however, mainly support attention allocation (see Figure 2.14). Drivers that are already focused on the driving scene are therefore less likely to benefit from additional knowledge about system fallibility. Future studies should therefore increase the duration of the experimental scenario to evoke complacency effects and decrease vigilance (see Subsection 2.6.2.2). Irrespective of the functional detail, the complete automated system was disengaged in case of (sub-)system failures. This could further strengthen trust contagion effects as it might lead users to assume that a failure of one subsystem extends to the whole system.

8.6 Summary and Conclusion

As highlighted by Lee and See (2004, p. 56), 'good calibration, high resolution, and high [temporal and functional] specificity of trust can mitigate misuse and disuse of automation'. Thus, (functional) specificity is one of three key components for supporting appropriate trust in automation (see Subsection 2.7.1.1). Similar to how the communication of system-wide uncertainties supported overall trust calibration (see Chapter 4), conveying the uncertainties regarding individual system functions increases the functional specificity of trust. Moreover, knowledge of function-specific uncertainties was shown to alleviate trust contagion effects (see Subsection 2.7.1.2 and Section 8.1). It was therefore hypothesised that increasing the functional detail of uncertainty information may have positive implications for human-automation interaction (see Subsection 2.9.3). This prompted an investigation with a driving simulator experiment, which was outlined in this chapter.

Contrary to expectations, the recorded takeover performance metrics did not indicate significant implications of function-specific uncertainty communication for driving safety. It can be argued that there may be a measurable impact of functional detail on takeover performance in scenarios that are longer and therefore more likely to evoke complacency effects (see Subsection 2.6.2.2) (Parasuraman and Manzey 2010).

Responses to semi-structured interview questions do, however, show that half of all participants value the additional detail and prefer to use the function-specific uncertainty display. The preference towards higher functional detail was mainly attributed to benefits regarding system awareness and attention allocation. In contrast, workload was a critical parameter for rejecting increases in functional detail.

The difference in subjective experience demonstrates that using the function-specific approach can positively impact driving experience despite a lack of measurable effects on driving performance data, warranting further exploration. Specifically, the display design should be iterated, with particular emphasis on factors identified in interviews. As such, particularly workload should be decreased while benefits to system awareness and attention allocation must be maintained.

The following chapter presents a design iteration that focuses on incorporating the presented factors. As implied by the variability of preference counts for each investigated display, users should be given the option to adjust the functional detail of uncertainty information. Therefore, the display should be designed in a way that allows users to reduce the functional detail towards system-wide uncertainty information if preferred.

CHAPTER 9

Visualisation of Function-Specific Uncertainties

Abstract Increasing the functional detail of uncertainty information was hypothesised to result in more appropriate trust and, consequently, improved human-automation interaction (see Subsection 2.9.3). While the results of the experiment outlined in Chapter 8 did not confirm the expected benefits for takeover performance metrics, interview responses showed that half of all participants preferred to use a display with higher functional detail, indicating positive implications for user experience. Particularly benefits to system awareness and attention allocation contributed to the reported preference indications, whereas workload was identified as a major inhibitor. As there was no consensus on preference, system designs should allow users to configure the functional detail. Based on these factors, an uncertainty display was developed that affords users the flexibility to vary the level of functional detail, whereby a particular emphasis was placed on reductions in workload. Augmented reality (AR) displays were preferred over alternative formats to reduce the cognitive effort required to link the conveyed information to reality. As such, a visual layout for an AR display is presented that allows the communication of lateral and longitudinal uncertainties. Using this layout, a set of 11 abstract visual variables was evaluated by 46 participants in terms of their ordinal characteristics and preference. In reference to preattentive properties (see Subsection 2.4.3), the variables were specifically selected to support rapid cognitive processing and thus reduce workload relative to more iconic representations (see Chapter 8). The results demonstrate that particularly changes in hue and pulse frequency exhibit guessable ordinal characteristics and are well-received. Aspects of this chapter, in part verbatim, are published in Kunze, Summerskill et al. (2018a, 2019c).

Background	Objectives	Method	Results	Implications
Existing research lacks recommendations for presenting function-specific uncertainties in the context of automated driving.	Develop a visual layout for a function-specific augmented reality uncertainty display and identify intuitive visual signifiers for conveying changes.	Mixed (within/ between) experimental design involving 46 participants that complete a forced choice sorting and a rating task in a laboratory setting.	Colour-based and animation-based variables exhibit both the most guessable ordinal characteristics and are well-received. No effect of driving function.	AR uncertainty displays should use changes in hue and pulse frequency to convey uncertainties. The results are valid irrespective of the driving function.

Figure 9.1: Overview of Chapter 9

9.1 Development of a Function-Specific Uncertainty Display

The previously outlined experiment (see Chapter 8) has highlighted a series of factors that are of paramount importance when designing function-specific uncertainty displays. As such, the work in Chapter 9 aims to create an increase in system awareness while minimising workload increments. Further, the display layout should be configurable to allow the communication of different levels of functional detail. To enable the comparison of several visualisation methods for the communication of uncertainty, a common foundation was developed to ensure internal validity (see Chapter 3).

9.1.1 Modality and Display Format

The development of a function-specific uncertainty display draws on prior research presented in Chapter 2. First, it must be determined which sensory channel should be addressed. Visual communication is best suited for conveying more complex information, such as function-specific uncertainties, without compromising on cognitive processing speed – which is crucial for critical information such as system uncertainties (see Table 2.1). Further, graphical representations of uncertainty have been shown to work equally well as numerical or verbal uncertainty indications (Bisantz, Marsiglio and Munch 2005) (see Subsection 2.8.2.2 and Subsection 2.9.3). Consequently, the visual channel is selected for investigation in the current application. The question remains which display format should be used to present the information, i. e. digital, analogue, representational, or contact-analogue displays (see Figure 2.10). The integration of uncertainties with the affected data was shown to result in more appropriate reliance (Neyedli, Hollands and Jamieson 2011). Further, contact-analogue displays require less cognitive effort to link the displayed information to reality (see Subsection 2.5.3.3 and Figure 2.10), potentially reducing workload while maintaining system awareness (see Section 8.4). Consequentially, the contact-analogue (or AR) display format was selected for the current application.

Finally, the visual method for conveying changes in uncertainty must be determined. Subsection 2.8.2.2 has highlighted that visual variables (see Figure 2.17) are well-suited for the (integrated) communication of uncertainties, particularly due to their preattentive properties (see Subsection 2.4.3). However, not all variables are appropriate for communicating an order. Specifically, the variables shape and arrangement do not possess sufficient ordinal characteristics and will therefore be excluded from further discussions (Roth 2017).

Importantly, the different connotation of these variables between their original context, maps, and the context of driving must be noted. Within GIScience, the conveyed uncertainties caution against the validity of underlying data. In the context of automated driving, the uncertainties are a measure of system confidence and, consequently, the likelihood of

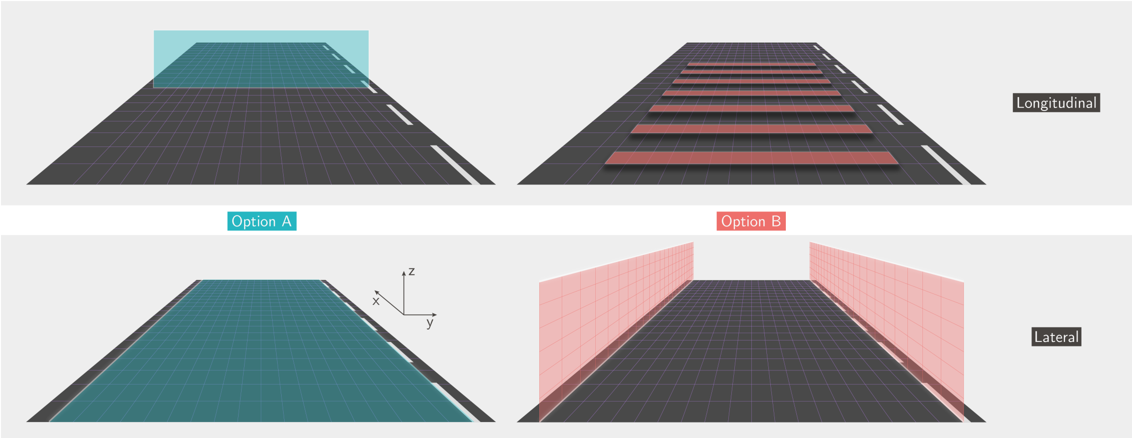


Figure 9.2: Basic layout options for presenting longitudinal (top) and lateral uncertainties (bottom) as overlays on AR displays (Kunze, Summerskill et al. 2018a)

takeovers. Whereas an increase in automation uncertainty implies an increased urgency to focus on the road, uncertainty in a geographic context may invoke the opposite reaction, i. e. a decreased focus on the areas affected by uncertainty.

9.1.2 Visual Layout

Taking a similar approach to that presented in Chapter 8, the DDT can be split into a lateral and a longitudinal component. Building on previously described work conducted within GIScience and the military domain, the uncertainty information can be superimposed on the FRD using visual variables. Overlays using visual variables lead, for instance, to a colouring of certain areas in the FRD. In a first step, this requires the definition of areas in the environment that will be overlaid. Derived from the movement of vehicles with the longitudinal axis x, the lateral axis y, and the transverse axis z, three planes are eligible: x-y, y-z, and x-z. As AR enables the contact-analogue visualisation of content, real-world analogies can be used to generate visual layout options. Regarding longitudinal vehicle motion, obstacles (see Figure 9.2, option A) or a series of bars (option B) analogue to the visualisation of ACC can be used as an overlay area. The road itself (option A) or crash barriers (option B) are possible analogies for lateral control. The overlay of lane markings was omitted as an option for this evaluation because it would not provide sufficient space for the visualisation of all visual variables. While this is also an issue for longitudinal option B, this was balanced with adjustments to the visualisation methods (see Subsection 9.2.2). Similar adaptations for the visualisation of lane markings were too subtle to notice.

Two visual layout options, one for each driving function, were selected based on their compatibility with each other. To allow for the function-specific communication of uncertainties, the overlays for each function must not interfere with each other. Using longitudinal and lateral option A, the overlays would overlap and may create ambiguity, particularly when colours blend into each other in semi-transparent designs. For similar reasons, the combination of longitudinal option B and lateral option A was excluded. While longitudinal option A and lateral option B would only slightly interfere on straight sections, the overlap would increase with curvature. As a consequence, a combination of lateral option B and longitudinal option B was selected as this was expected to result in unambiguous visualisations with minimal overlaps.

9.2 Method

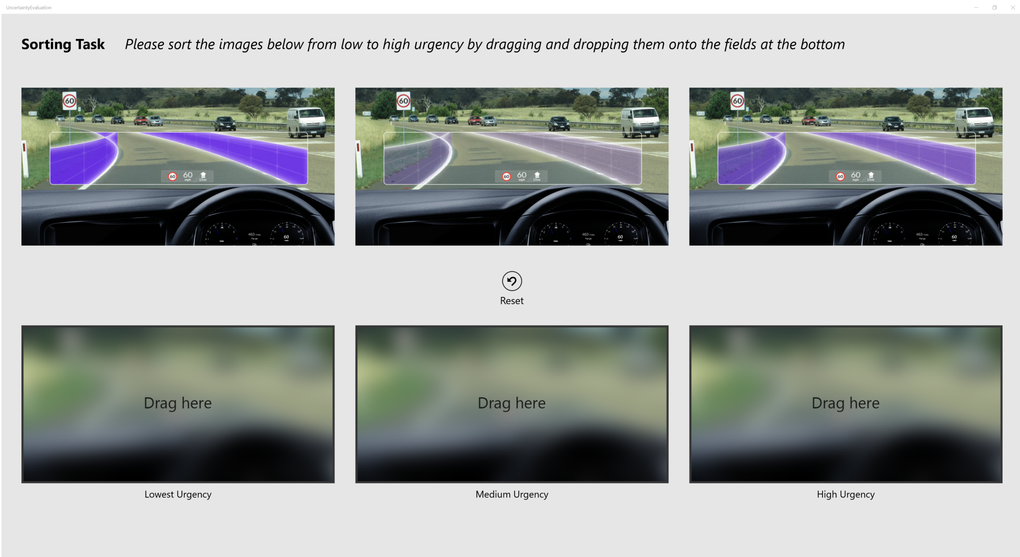
Using the developed function-specific uncertainty display, the following research questions were addressed:

1. Which visual variables are best suited for conveying a guessable order for changes in uncertainty?
2. Which visual variables are best received for the purpose of conveying uncertainties?
3. How do preference scores vary with respect to each investigated (driving) function?

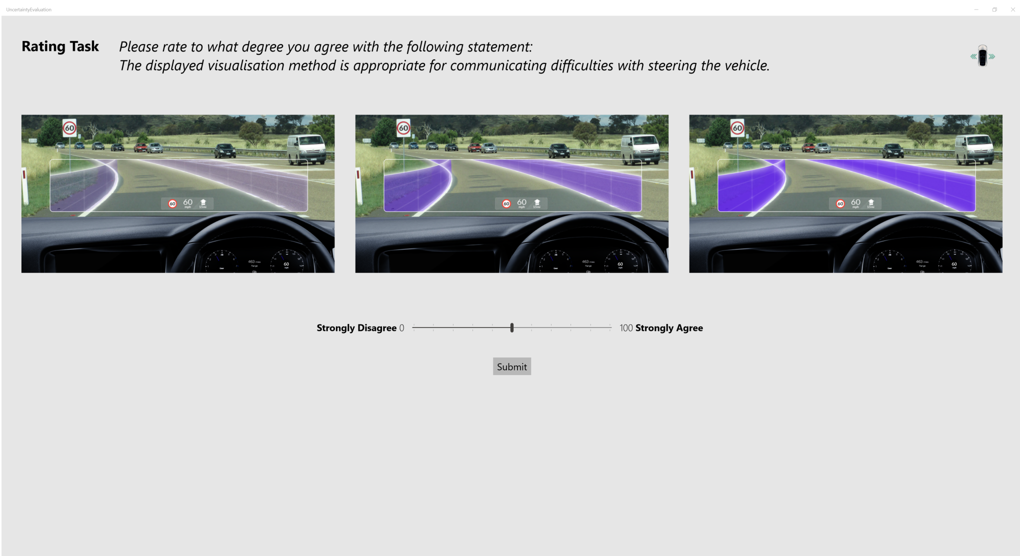
A laboratory-based experiment using a series of tasks presented on a computer screen was conducted to evaluate a large set of basic visual variables regarding their suitability for communicating lateral and longitudinal uncertainties using an AR display. In reference to Jordan (1998), the suitability is thereby determined by investigating the guessability of different variables and subjective preference ratings. The results of this study enable the informed development of future AR (uncertainty) displays and significantly reduce the variables that are to be considered in future investigations.

9.2.1 Participants

The experiment was completed by 46 participants (16 female) with an average age of 30.239 years ($SD = 9.403$), ranging from 18 to 54 years. Participants estimated to drive an average of 10,338.040 ($SD = 28,999.640$) miles annually and reported to hold a driving licence for an average of 10.109 years ($SD = 9.137$). Five participants had experience with driving assistance systems such as ACC or lane keeping assists. Participants received no monetary compensation for taking part in the study, ethical clearance was granted by the Ethics Committee of Loughborough University.



(a)



(b)

Figure 9.3: Screenshots of the interface layout for sorting (a) and rating session (b) (Kunze, Summerskill et al. 2018a)

9.2.2 Uncertainty Visualisation Method

The uncertainty visualisation method applied in the current study follows the fundamental display concept discussed in Section 9.1. The selection of visual variables for this experiment was performed based on anticipated ordinal characteristics (Roth 2017) and their practicability in the given context. A total of 11 visual variables were selected, specifically variables a to n as shown in Figure 2.17 with the exception of arrangement, shape, and resolution.

In conformity with MacEachren, Roth et al. (2012), three instances corresponding to low, medium, and high uncertainties were designed for each variable and driving function, leading to a total of 66 instances. To contextualise the variables, each of these instances was rendered into a generic road environment (see Figures 9.3 and 9.4).

The specific implementations of the variables were aimed to be similar for both driving functions. However, several differences were required to ensure the visibility of more subtle variables when using the longitudinal layout. As such, line orientation and grain were not implemented as a fill pattern but rather the bars themselves were used to convey changes (see Figure 9.4). Moreover, position was varied on the y-axis for lateral uncertainties to convey the narrowing of lanes whereas the z-axis was used for longitudinal uncertainties to signalise barriers. The instances of animation-based variables differed in terms of the flashing rate or movement speed. Movement was created by animating the position of the elements. The distinguishability of the variable instances was ensured through a pilot study.

9.2.3 Design

The experiment employed a 2 (driving functions, between subjects) × 11 (visual variables, within subjects) mixed design to explore the ordinal characteristics of visual variables and their suitability for conveying uncertainties in the context of automated driving. To prevent order effects, the order of the levels in which the visual variables were displayed was randomised.

9.2.4 Apparatus

The experimental tasks were implemented in a purpose-built Universal Windows Platform application to enable the accurate measurement of response times and ensure similar conditions for all participants. To prevent effects of different display specifications (resolution, size, etc.) and calibrations (contrast, brightness, colour, etc.), participants were required to complete the experiment in a laboratory using a 27" monitor with a resolution of 3840 × 2160 pixels and fixed calibration settings. Moreover, both lighting and sound levels were controlled to improve internal validity. Throughout the experiment, participants used a computer mouse and a keyboard to provide input and complete the tasks. All required information was integrated in the application as animated sequences with voice-overs. To minimise distress and answer potential questions, an experimenter and a chaperone were present for all sessions.

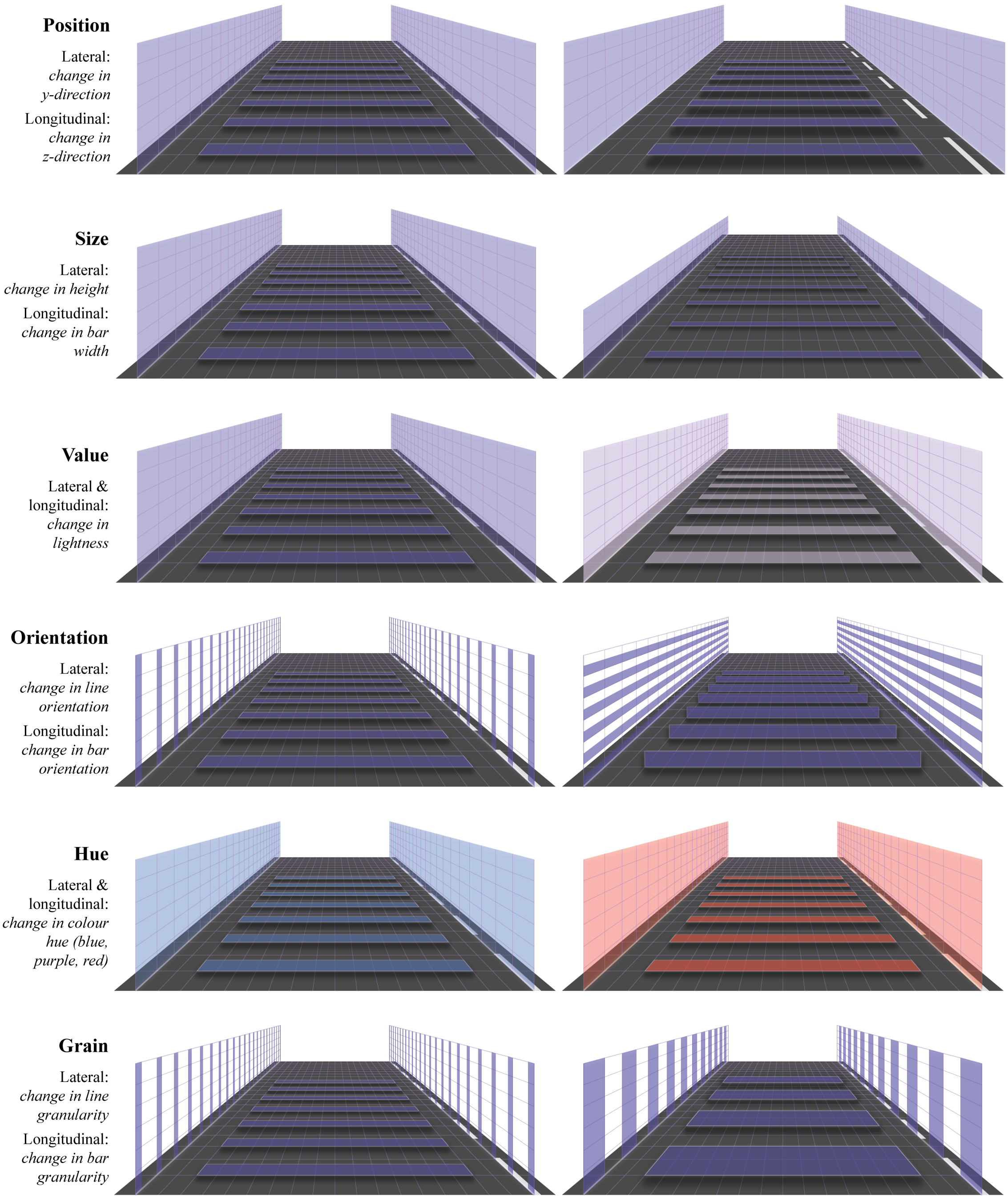


Figure 9.4: Implementation of visual variables in the chosen layout options, ordered according to condition 1 with increasing uncertainty from left to right; intermediate steps were omitted to improve visibility in print (see Figure 2.17) (Kunze, Summerskill et al. 2018a)

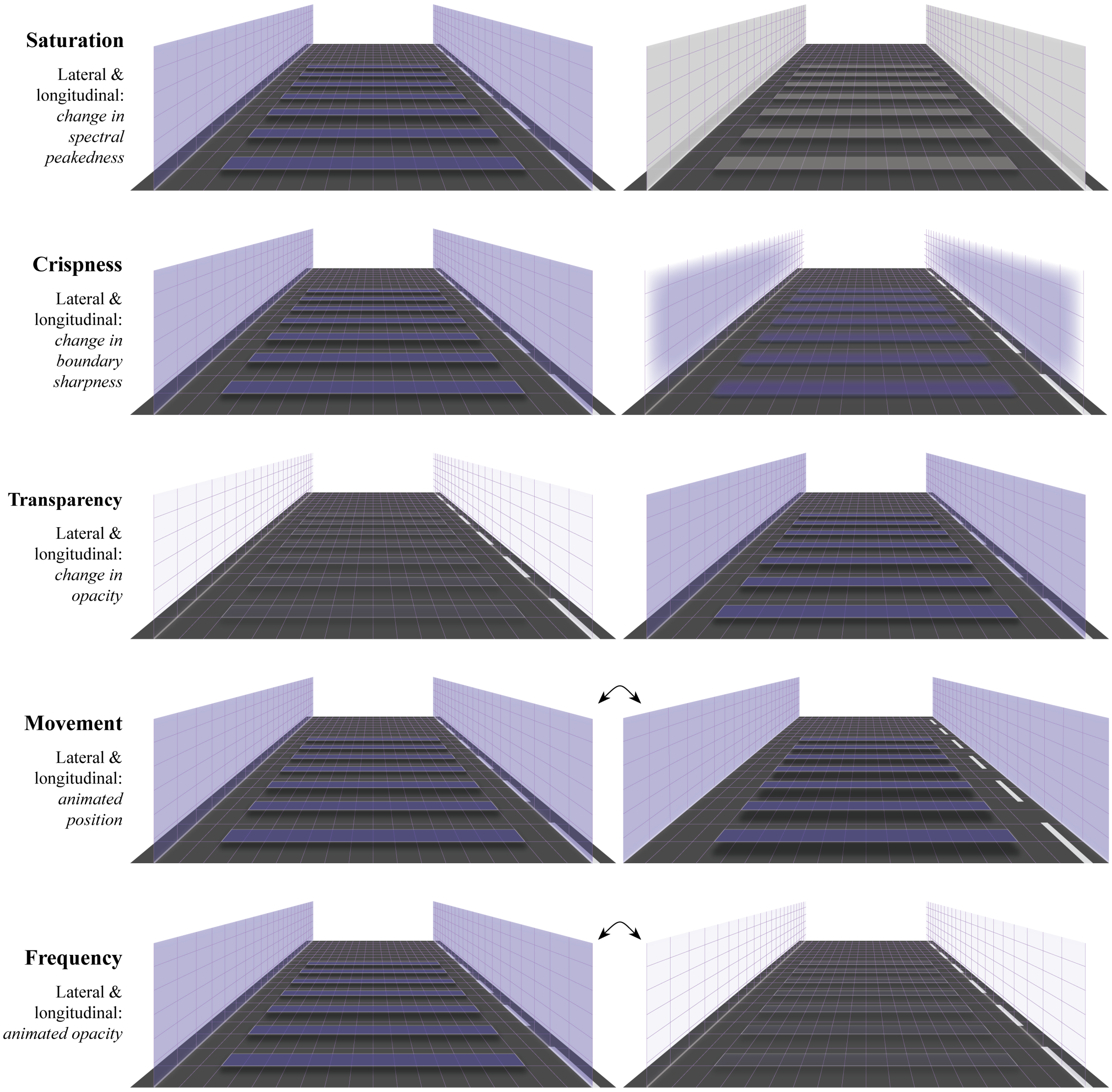


Figure 9.4: (Continued) Implementation of visual variables in the chosen layout options, ordered according to condition 1 with increasing uncertainty from left to right; intermediate steps were omitted to improve visibility in print (see Figure 2.17) (Kunze, Summerskill et al. 2018a)

9.2.5 Procedure and Scenario

The research questions were investigated in two consecutive sessions with a total duration of approximately 30 minutes. The order of these sessions was fixed, with the second session building on information of the first. Purpose and instructions regarding each session were conveyed using videos. Following the completion of the introduction, participants were given the opportunity to ask questions. The comprehension of each task was assessed with a simplified practice question. Individual tasks were separated by an info screen that indicated the number of remaining tasks and allowed for a brief mental break. The controlled repositioning of the mouse cursor before each task was ensured through the fixed position of a button linked to the subsequent task.

The ordinal characteristics of each variable were assessed in the first session. Specifically, the sorting consistency between participants was investigated using a series of 11 forced choice sorting tasks, one for every visual variable (see Figure 9.3a). To prevent order effects, the sequence of these tasks was randomised. Using a drag-and-drop interaction, participants had to arrange three randomly ordered variable instances in terms of their perceived urgency. As a consequence of ambiguities observed during pilot studies, the term uncertainty was substituted with urgency to improve clarity. In the context of driving automation, urgency and uncertainty both imply a higher likelihood of takeovers. The response time for each task was recorded as an additional measure for the ordinal characteristics of each variable. The quicker variable instances were sorted, the more guessable and clear was the order assumed to be (MacEachren, Roth et al. 2012). Response time was measured in milliseconds between the click of the proceed button on the info screen and the completion of the sorting task.

The preference levels for each variable concerning the communication of lateral/longitudinal uncertainties was assessed in the second session. Despite having clear ordinal characteristics, variables may not be suited for uncertainty communication. As shown in Figure 9.3b, participants were again presented with three instances of each visual variable. In contrast to the firsts session, the order was not randomised but replicated the sequence indicated by the participant during the first session. To improve clarity, the terms lateral and longitudinal were replaced by the expressions *steering the vehicle* and *braking and accelerating.* Moreover, a graphical representation positioned in the top right corner clarified the driving function. Participants reported their preference for each variable using a slider scale ranging from *strongly disagree* (0) to *strongly agree* (100). A slider scale was preferred over a Likert scale to increase granularity. The use of pairwise comparisons was avoided to reduce mental strain on participants. In contrast to the 11 questions required for the chosen implementation, a pairwise comparison would have required a total of $\frac{N \cdot (N-1)}{2} = 55$ questions.

9.2.6 Data Analysis

The key objective of the sorting task was to investigate the ordinal characteristics of the visual variables. To facilitate the evaluation of the sorting task results, a sorting consistency measure for ordinal scales was developed. In comparison with related work that used six individual values to characterise sorting consistency (Noah, Gable and Walker 2016), the use of a single measure improves the comparability between variables.

The following fundamental conditions provide the basis for the sorting consistency measure:

$$\begin{array}{ll} \text{Condition 1:} & \text{Extreme A} < \text{Intermediate} < \text{Extreme B} \\ \text{Condition 2:} & \text{Extreme B} < \text{Intermediate} < \text{Extreme A} \end{array}$$

The conditions are based around the assumption that all selected variables have sufficient ordinal characteristics (Roth 2017), leading to a general order from one extreme to another with a fixed intermediate value at the centre. It must be investigated, however, which extreme conveys the lowest (and which the highest) urgency. For instance, does an increase in transparency lead to a perceived increase in uncertainty? Thereby, condition 1 refers to the order indicated by previous publications, as shown in Figures 2.17 and 9.4, whereas condition 2 inverts the order. The degree of fulfilment for each condition can be quantified through the evaluation of their three constituent requirements.

To achieve a complete degree fulfilment, i. e. 100 %, for condition 1, Extreme A has to be placed left of the Intermediate and both Extreme A and the Intermediate have to be placed left of Extreme B (see Figure 9.3a). The fulfilment of zero requirements leads to a score of 0 % while the fulfilment of one or two requirements leads to 33.3 % and 66.7 % respectively. Combined, the degrees of fulfilment for each condition add up to 100 %. The higher score indicates the more logical order as perceived by each participant.

Based on these conditions, the sorting scores for each participant and visual variable were recorded. The average score within each group for a each condition then serves as a measure for sorting consistency, whereas the higher scoring condition indicates the sorting direction. For instance, the average sorting score of a variable may be 0.7 for condition 1 and 0.3 for condition 2. Consequently, condition 1 is assumed to be the more guessable order. Higher scores indicate a higher sorting consistency, whereby 0.5 indicates the minimum score.

9.3 Results

The responses of participants were collected and processed by the purpose-built application and subsequently analysed in R.

Table 9.1: Sorting scores for each variable and driving function for the higher scoring condition; *Cond.*: sorting condition with higher mean score, condition 1 refers to the order shown in Figure 9.4, condition 2 means that the order shown in the figure is reversed (Kunze, Summerskill et al. 2018a)

Visual	**Lateral**				**Longitudinal**			
Variables	*Cond.*	*Mean*	*Mode*	*SD*	*Cond.*	*Mean*	*Mode*	*SD*
Position	1	0.565	1.000	0.420	1	0.826	1.000	0.316
Size	2	0.754	1.000	0.417	2	0.913	1.000	0.251
Value	2	0.884	1.000	0.312	2	0.942	1.000	0.217
Orientation	1	0.812	1.000	0.281	1	0.884	1.000	0.238
Hue	1	0.957	1.000	0.115	1	0.986	1.000	0.070
Grain	1	0.754	1.000	0.405	1	0.522	1.000	0.491
Saturation	2	0.957	1.000	0.209	2	0.928	1.000	0.245
Crispness	2	0.768	1.000	0.420	2	0.594	1.000	0.426
Transparency	1	0.913	1.000	0.288	1	0.870	1.000	0.344
Movement	1	0.913	1.000	0.251	1	0.710	1.000	0.338
Frequency	1	0.899	1.000	0.274	1	0.899	1.000	0.234

9.3.1 Sorting Scores

The results of the forced choice sorting task are summarised in Table 9.1. Irrespective of variable or driving function, the mode took extremes values, i. e. 1.00, for one of the sorting conditions. Hence, the intermediate instance was generally placed between two extremes, affirming the assumption that all variables conveyed an order corresponding to one of the sorting conditions. The dominant sorting condition was consistent between groups. In contrast to previous research within GIScience (MacEachren, Roth et al. 2012), however, the sorting direction of size, value, saturation, and crispness was reversed (see Figure 2.17). Indicative of a high sorting consistency, the variable hue received the highest sorting mean scores for both the lateral and longitudinal driving function, with 0.957 and 0.986 respectively. This is reaffirmed by the lowest standard deviation, suggesting that few participants chose a different sorting order. Besides hue, the sorting scores indicate a high sorting consistency for the variables saturation, transparency, frequency, and value.

9.3.2 Response Times

There were no statistically significant differences in response time between or within group as determined by a 2×11 mixed design ANOVA. However, a significant interaction between the independent variables was returned, $F(1, 10) = 1.989, p = .033$. This indicates that response times for individual variables differed between groups.

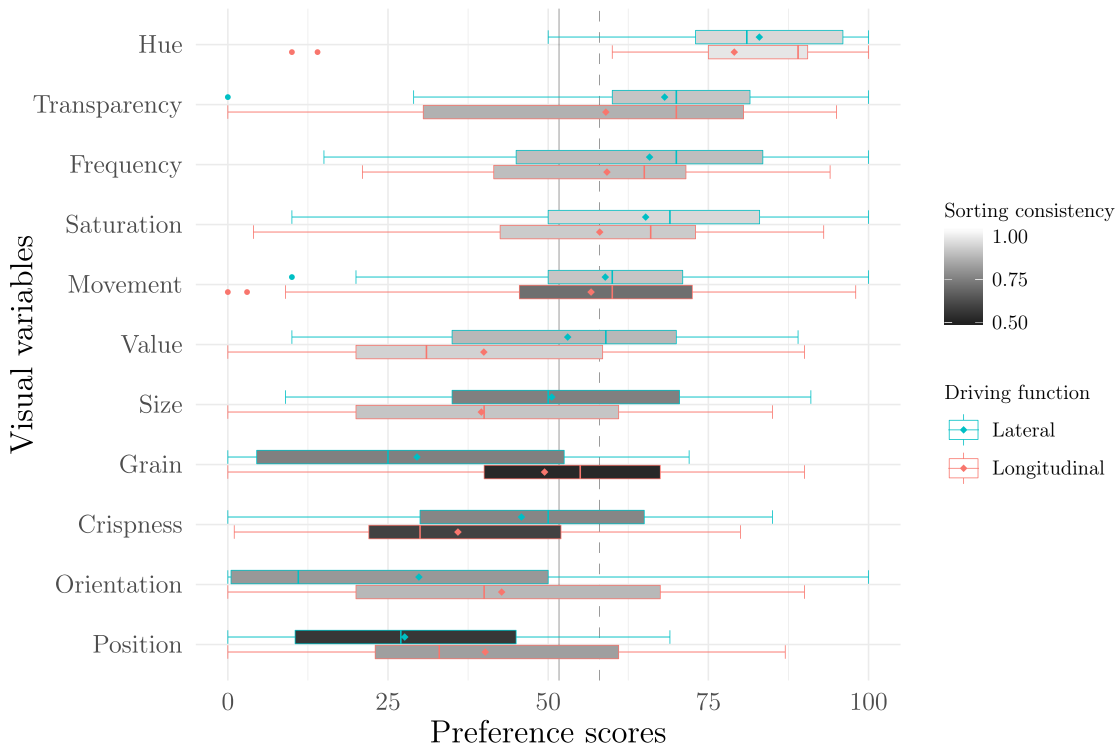


Figure 9.5: Preference scores for each visual variable relative to the driving function. Diamonds indicate the mean for each variable. The dashed line highlights the median, the full line the mean of the overall preference scores (Kunze, Summerskill et al. 2018a).

9.3.3 Preference Scores

Both a Shapiro-Wilk test and the visual inspection of q-q plots indicated that the preference scores were not normally distributed. Consequently, a nonparametric rank-based test equivalent to a 2 × 11 mixed design ANOVA was performed (Feys 2016; Noguchi et al. 2012). The mean preference scores between visual variables differed, as evidenced by a significant main effect of visual variables on preference scores, $F(1.000, 6.652) = 20.020, p < .001$. Moreover, a significant interaction was returned, $F(1.000, 6.652) = 2.561, p = .014$. The preference scores are visualised in Figure 9.5. The visual variable hue returned the highest preference ratings for both the lateral ($M_{lat} = 82.957$, $SD_{lat} = 13.465$) and the longitudinal group ($M_{long} = 79.043$, $SD_{long} = 23.679$), followed by transparency ($M_{lat} = 68.174$, $SD_{lat} = 22.413$, $M_{long} = 59.000$, $SD_{long} = 28.358$), frequency ($M_{lat} = 65.826$, $SD_{lat} = 24.304$, $M_{long} = 59.174$, $SD_{long} = 20.198$), and saturation ($M_{lat} = 65.217$, $SD_{lat} = 25.717$, $M_{long} = 58.043$, $SD_{long} = 25.829$). Fdr-corrected (Benjamini and Hochberg 1995) post-hoc pairwise sign tests were conducted to further assess the differences. Particularly the variable hue differed significantly from the remaining variables (see Table 9.2).

Table 9.2: Significant p values for fdr-corrected post-hoc pairwise sign tests on preference scores (above bullets: longitudinal; below: lateral) (Kunze, Summerskill et al. 2018a)

	Position	Size	Value	Orientation	Hue	Grain	Saturation	Crispness	Transparency	Movement	Frequency
Position	•	-	-	-	$< .001$	-	-	-	-	-	-
Size	.044	•	-	-	$< .001$	-	-	-	-	-	-
Value	.005	-	•	-	$< .001$	-	-	-	.009	-	-
Orientation.	-	.012	.012	•	.003	-	-	-	-	-	-
Hue	$< .001$	$< .001$	.001	$< .001$	•	$< .001$	$< .001$	$< .001$	.002	.020	.020
Grain	-	-	.023	-	$< .001$	•	-	-	-	-	-
Saturation	.005	.031	.023	.023	-	.016	•	.014	-	-	-
Crispness	-	-	-	-	.001	-	.034	•	-	-	-
Transparency	.004	.012	.023	.016	.016	.004	-	.009	•	-	-
Movement	.012	-	-	.031	.003	.004	-	-	.044	•	-
Frequency	.004	-	-	.005	.016	.004	-	-	-	-	•

9.3.4 Response Times and Preference Scores

Multiple regression analysis was employed to account for the repeated measures design and, consequently, non-independent data points. Specifically, it was assessed if response times predicted preference scores under consideration of the factor subjects (Bland 2015).

The results of the regression analysis suggest that the predictors explained 22.9 % of the variance in the lateral group ($R^2 = .229, F(23, 229) = 2.857, p < .001$) and 17.2 % in the longitudinal group ($R^2 = .172, F(23, 229) = 2.062, p = .004$). It was found that response time predicted preference significantly in the lateral group ($\beta = -1.545, p < .001$) and in the longitudinal group ($\beta = -0.697, p = .041$).

9.3.5 Response Times and Sorting Scores

Adopting a similar rationale as in the previously outlined statistical evaluation, multiple regression analysis was used to investigate if the response times significantly predicted sorting scores, taking out subjects as a factor (Bland 2015).

The results of the regression analysis indicate that the predictors explained 14.9 % of the variance in the lateral group ($R^2 = .149, F(23, 229) = 1.745, p = .022$) and 15.6 % in the longitudinal group ($R^2 = .156, F(23, 229) = 1.835, p = .014$). It was found that response time predicted sorting scores significantly in the lateral group ($\beta = -0.239, p = .043$) and in the longitudinal group ($\beta = -0.412, p = .006$).

9.4 Discussion

The experimental results show a series of implications for the communication of function-specific uncertainties using AR displays.

To fulfil the requirements deducted from the interview responses in Chapter 8, namely improving system awareness while reducing workload increments, some variables should be preferred over others as they are characterised by a more guessable order. Using variables which are (a) characterised by a guessable order and (b) considered appropriate for uncertainty communication likely promotes the use of intuitive cognition, leading to faster, unconscious responses that require less cognitive effort (see Subsection 2.4.3). Within the current study, the faster processing is indicated by a significant prediction of sorting scores by response time.

The results of the forced choice sorting task show that hue, saturation, transparency, frequency, and value were ordered most consistently and hence possess the strongest ordinal characteristics (research question 1, see Section 9.2). The higher scoring sorting condition was thereby consistent between driving functions, indicating that the order direction was independent of the chosen implementation. The results shown in Table 9.1 further highlight that the order direction was reversed when compared with the findings of MacEachren, Roth et al. (2012). This can be attributed to the aforementioned differences in connotation between automated driving and GIScience.

Further, the visual variables differed in terms of preference, as indicated by a main effect of visual variables on preference scores (research question 2, see Section 9.2). The results of post-hoc pairwise sign tests suggest that especially changes in hue are well-received for both lateral and longitudinal uncertainty communication. This may be a consequence of the strong preattentive effects associated with colour hue (Ware 2013). Moreover, the use of colour hue to indicate changes in urgency is likely more familiar to participants, leading to a preference due to a mere exposure effect. The significant prediction of preference ratings by response times affirms this assumption. In other words, variables with more readily identifiable ordinal characteristics received, on average, higher preference scores. The high ratings for colour contrast the findings in the context of GIScience. This can be attributed to the less familiar colour hue instances (olive, green, purple) used by MacEachren, Roth et al. (2012).

Statistically non-significant main effects of the driving function on both preference scores and response times suggest that the chosen visual variable implementations work, overall, equally well for conveying both lateral and longitudinal uncertainties (research question 3, see Section 9.2). Thus, the identified variables for research questions 1 and 2 would likely be suitable in comparable implementations, for instance when only system-wide uncertainties are conveyed as a generic road overlay (see lateral option A, Figure 9.2). Figure 9.6 shows a possible implementation based on the discussed results. Changes in colour hue from blue to red can be used to convey increases in uncertainty. To indicate increased urgency for the highest uncertainty level, animations corresponding to the variable frequency were added.

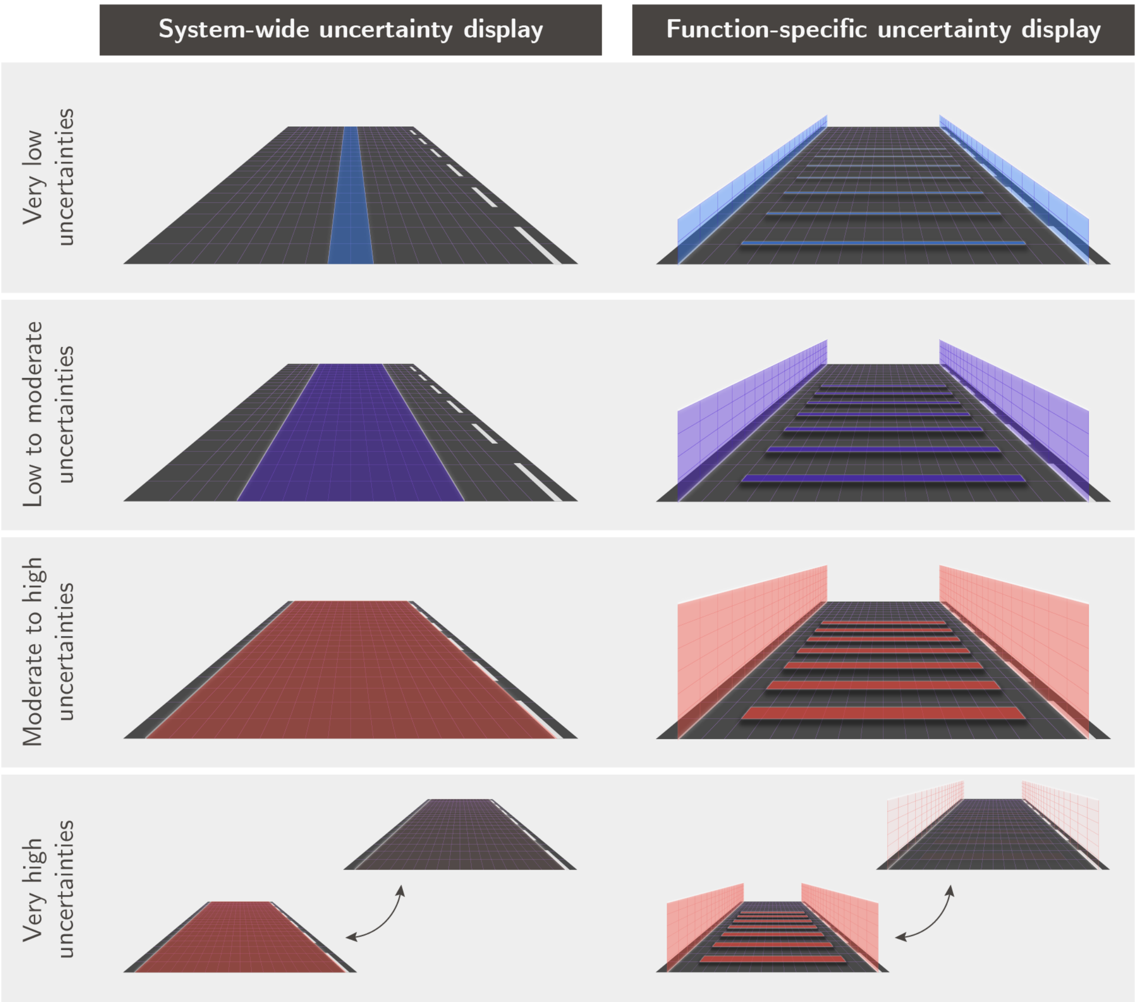


Figure 9.6: Possible implementation of an AR uncertainty display based on experimental results. Uncertainty increases from top to bottom in reference to the levels specified in Table 6.11. The system-wide uncertainty display uses a generic layout that overlays content onto the road. The function-specific uses the layout specified in Subsection 9.2.2. The function-specific uncertainty indications for lateral and longitudinal vehicle motion can change independently from each other.

For reasons of accessibility and processing speed Subsection 2.4.3), the information should be redundantly coded. As the colour-related variables transparency and value would reduce the visibility of hue, the variable size was added.

Given that there was a significant interaction for the prediction of response times and preference scores, the suitability of some variables depends on the driving function. This can be attributed to differences in implementation for those variables that did not rely on fill colour or alpha values. The variable position, for instance, was varied along the y-direction to indicate lateral uncertainties whereas a change in z-direction was used to convey longitudinal uncertainties. In an analogous manner, the implementations of the variables grain and orientation differed. These constraints do not apply to the variables identified in response to research questions 1 and 2 (see Section 9.2).

9.5 Limitations

The experimental results are confined by the following limitations. Rendering the visual variables as part of a driving scene induces several degrees of freedom in contrast to the evaluation of abstract features in isolation. This includes, for instance, the definition of a default colour, the transfer of abstract variables to the AR layout, the design of specific variable instances, and the selection of a driving scene. Moreover, differences in cultural background may have an effect on the dependent variables evaluated in this experiment. Colour hues, for example, convey a different meaning depending on the cultural context (Heimgärtner, Solanki and Windl 2017).

Additionally, future investigations should evaluate the visual variables in a dynamic driving context to assess aspects such as occlusion and to ensure that the changes in urgency are perceptible without seeing the variable instances next to each other (see Figure 9.3a).

Moreover, potential limitations concerning the practical implementation of AR interfaces must be noted. In particular, the accurate positioning of contact-analogue information will be challenging to achieve. Further, the interaction between synthetic light originating from the HUD and real-world light with dynamically changing contrasts likely impacts the visibility of AR content (Gabbard, Fitch and Kim 2014). As such, some visual variables, for instance transparency, may be rendered unsuitable in the actual usage environment. Nevertheless, the presented results reduce the number of visual variables that can be considered suitable for AR-based uncertainty communication.

9.6 Summary and Conclusion

Grounded in the recommendations for function-specific uncertainty communication established in Chapter 8 and research concerning human information processing (see Section 2.4), a display design for conveying uncertainties varying in functional detail was developed. To reduce the cognitive effort required to link the conveyed information to the real-world environment, contact-analogue (AR) displays were used (see Figure 2.10). For this purpose, a visual layout was developed to accommodate various visual variables that promote fast cognitive processing and thereby reduce workload (see Subsection 2.4.3). The results of a sorting task indicate that particularly changes in hue and pulse frequency have a more guessable order and are preferred by participants. The highest scoring variables were equal for both evaluated system functions, suggesting that the results are valid for various implementations using AR displays. As such, the results do not only apply to function-specific but also system-wide uncertainty displays (see Figure 9.6 for a possible implementation).

While the presented results can inform the design of AR displays, more research is needed to assess the impact of function-specific, AR-based uncertainty communication on driving-related measures and workload, potentially in the context of partial takeovers.

CHAPTER 10

Discussion and Conclusion

This research project investigated uncertainty communication as a means for enhancing user experience and safety in the context of automated driving. Using a controlled stepwise approach, various aspects of uncertainty communication were analysed in terms of their effects on driving performance metrics and human factors (HF)-related variables such as workload, trust, and situation awareness (SA). This chapter consolidates and critically examines the experimental findings and derived contributions of this research project. Based on the discussed results and limitations, an outlook for future work is provided and an overall conclusion is given.

10.1 Broader Relevance of Experimental Results

The elimination of road traffic crashes attributable to human error is one of the major promises of driving automation. As long as humans are involved in the driving task, even if only temporarily or as a fallback measure, the complete elimination of human-related causes cannot be accomplished. Consequently, developers strive to achieve full driving automation, i. e. SAE level 5 (see Table 2.3).

The Organisation for Economic Co-operation and Development (OECD) identified two incremental development pathways towards SAE level 5 (OECD 2015). Traditional car manufacturers are likely to embrace a strategy that aims to gradually increase the level of driving automation. Targeted at customers seeking private ownership, the human remains to be involved in the driving task for at least parts of the journey until level 5 is achieved. This strategy can be referred to as *something everywhere* (OECD 2015, p. 13). Contrasting this approach, developers that regard mobility as a service aim to first deploy vehicles without human drivers for a limited operational domain and later extend it to other usage contexts. This approach can be labelled as *everything somewhere* (OECD 2015, p. 13).

The findings of this research project are of particular relevance for the former pathway towards full driving automation. The significance of uncertainty communication for the *something everywhere* strategy stems from a combination of issues pertaining to human factors and legal responsibility. For SAE levels 1 to 4, users are likely to be involved in the driving task for at least parts of a journey. Knowledge of system uncertainties enables users to make informed judgements about the appropriateness of ceding control to the automated driving system. Without a means for judging the time-specific and context-dependent competence of the automated driving system, users are likely taken by surprise in the event of system failures. Lawmakers have acknowledged the need of avoiding *black box* AI-based systems and call for an increase in transparency to make AI more understandable (House of Lords Select Committee on Artificial Intelligence 2018). Time-specific and context-dependent knowledge of uncertainties allows users to gradually learn about system limitations, thus promoting the build-up of an appropriate mental model. As detailed in the following section, uncertainty communication significantly benefits key parameters pertaining to human-automation interaction, such as SA, trust, and takeover performance. Uncertainty displays can therefore be regarded as a safety feature for all automated driving systems that have the option to be driven manually, i. e. SAE levels 1 to 4. Moreover, uncertainty communication can enhance user experience by supporting the guilt-free engagement in NDRTs. Regulators emphasise that, at least for the near future, the principle that engagement in NDRTs should be minimised is to remain true (UNECE 2017). Knowledge of uncertainties allows users to distinguish between situations in which it is safe to engage in NDRTs and those that demand their attention on the road.

Besides addressing HF-related issues, uncertainty communication may also reduce ambiguities concerning legal responsibilities. For instance, the degree of system uncertainty can serve as an indicator for the required monitoring behaviour of the human operator during activated driving automation. As regulators are currently considering mandatory driver monitoring systems to assess user readiness and awareness prior to takeovers (UNECE 2019), knowledge of uncertainties can support users in determining behaviour that is conforming to the law. In fact, the responsibilities of users during activated driving automation remain to be specified by national regulators. In a consultation paper requested by the Centre for Connected and Autonomous Vehicles, a high profile group of the UK government, the Law Commission of England and Wales and the Scottish Law Commission proposed the category *user-in-charge* for operators of SAE level 4 systems. As user-in-charge, operators are required to remain 'qualified and fit to drive (Law Commission of England and Wales and Scottish Law Commission 2018, p. 36)' during activated driving automation in order to be prepared for planned takeovers or the continuation of the journey after the automated driving system has achieved a minimal risk condition. However, they are 'entitled to undertake secondary activities (Law Commission of England and Wales and Scottish Law Commission 2018, p. 38)', which may be further regulated (UNECE 2017). While the consultation states the principle that the user-in-charge would not be held responsible for the driving behaviour in case of criminal offences, questions arise if there should be exceptions provided that the risks were obvious (Law Commission of England and Wales and Scottish Law Commission 2018, p. 41). As discussed in Subsection 2.6.2.1, automated systems are inherently brittle and may unexpectedly fail due to exceeded system limitations. As a consequence, unexpected changes in the driving environment, for instance due to heavy fog, may prepone planned takeovers. An uncertainty display could convey the risks resulting from unexpected environmental factors and support the driver with re-engaging in the driving task if required. Further, it would be naïve and negligent to assume that users of automated driving systems could – and, in fact, would be willing to – memorise limitations regarding operational domains, particularly when considering the frequency of software updates and trends such as shared mobility. The communication of uncertainties provides a means to dynamically convey system capabilities in a more general, time-specific, and context-dependent manner that does not rely on the correct memorisation of all system limitations.

Combined, the presented examples concerning human factors and liability highlight that the topic of uncertainty communication is likely to remain relevant for both developers and regulators. The experimental results presented in this thesis can be regarded as guidelines for the practical implementation of user interfaces for uncertainty communication.

10.2 Discussion of Experimental Results

The experimental research presented in this thesis focused on vehicles equipped with conditionally automated driving systems, corresponding to SAE level 3 (see Table 2.3). While such systems are capable of performing the complete DDT on a sustained basis (SAE International 2018), drivers are required to remain fallback-ready in case of system failures or an exceedance of system limitations. While this affords drivers the flexibility to engage in NDRTs (see Subsection 2.5.2), users are completely removed from the feedback loop of driving, leading to a series of HF challenges. As drivers are likely not monitoring the FRD while the automation is engaged (see Subsection 2.5.2), awareness of their surroundings, i. e. SA, is reduced, thus impeding safe takeovers in case of system failures (Endsley 2017; Endsley and Kiris 1995). Inappropriate trust in the capabilities of the automated system can be regarded as a major cause for this behaviour (Lee and See 2004; Hoff and Bashir 2015). To facilitate more appropriate trust, operators can be made aware of system limitations through uncertainty communication (see Section 2.8). As predicted by the integral model presented in Figure 2.14, time-specific knowledge of system uncertainties was expected to support operators in establishing appropriate trust levels that match the true capabilities of an automated system. As trust in automation and monitoring behaviour were shown to be linked (Hergeth et al. 2016; Muir and Moray 1996), improved trust calibration would lead to adapted monitoring behaviour that consequentially results in improved SA and takeover performance.

The experiments outlined in Chapters 4 to 7 confirm the predicted benefits of uncertainty communication for trust calibration. Specifically, the recorded eye tracking and NDRT performance data as well as questionnaire responses unanimously evidence that uncertainty communication systematically affects operator trust. The results also confirm the previously identified relations between trust and monitoring behaviour (Hergeth et al. 2016; Muir and Moray 1996). In agreement with closely related work and as predicted by the integral model (Beller, Heesen and Vollrath 2013; Helldin et al. 2013), the experimental results further show improvements for takeover performance metrics. In combination, the results confirm the predicted series of reactions evoked by uncertainty communication (see Figure 2.14). As such, uncertainty communication leads to more appropriate trust which, in turn, results in more appropriate top-down attention allocation. The optimised allocation of attentional resources leads to a higher focus on the FRD prior to critical situations, resulting in improved SA and takeover performance.

Besides general implications of uncertainty communication for human-automation interaction, the experiments further provide insights regarding appropriate system designs. As discussed in Section 2.4, human information processing is delimited by finite attentional resources, prompting the need for stimuli that are effortlessly processed with few attentional demands and address unoccupied resource pools (Wickens, Hollands et al. 2013). The experiments outlined in Chapters 6 to 9 confirm the anticipated significance of attention and workload in the context of human-automation interaction. Specifically, reductions in information access effort and the use of unoccupied sensory channels resulted in improved subjective user experience (UX) scores and workload indications as well as safer takeovers (see Chapters 6 and 7). Further, workload was a contributing factor for favouring lower levels of functional detail (see Chapter 8). In the experiment outlined in Chapter 9, participants preferred visual variables that were more quickly sorted into the corresponding categories, indicating the perceptible advantages of effortlessly processed variables. As the benefits were observed across different modalities, designers are encouraged to select unoccupied sensory channels for conveying uncertainties depending on the context of use (Wickens 2008). In the following, the experimental results are discussed in regard to each corresponding research question (see Table 1.1).

10.2.1 Research Question 1: What are the implications of dynamic uncertainty communication on key parameters of human-automation interaction?

Existing research has highlighted an interrelation among several factors pertaining to human-automation interaction, foremostly SA, workload, and trust (see Figure 2.14). While previous work has indicated the general benefits of uncertainty communication for a subset of these factors (Beller, Heesen and Vollrath 2013; Helldin et al. 2013), no publications known to the author have addressed the interplay of these factors comprehensively. Further, the existing body of research lacks an assessment of the dynamic fluctuations of these factors depending on different uncertainty levels (see Subsection 2.9.1).

Addressing these shortfalls and expanding on previous work regarding uncertainty communication, a driving simulator experiment was conducted. To allow a comparison with existing work (Beller, Heesen and Vollrath 2013; Helldin et al. 2013), the uncertainties were conveyed using a conventional digital instrument cluster display. For the assessment of dynamic changes in dependent variables, an uncertainty display with multiple levels was designed, i. e. corresponding to very low, low to moderate, moderate to high, and very high uncertainties. As human-like features were shown to promote trust in automation (see Subsection 2.8.2.3) and were found intuitive in a related study (Beller, Heesen and Vollrath 2013), anthropomorphic features were used to convey the uncertainty information (see Figures 4.2 and 10.1). To replicate realistic usage scenarios, participants were engaged in a visual search task while the automation was activated. Moreover, the system failed only on a single occasion to account for the high reliability levels that commercially available automated driving systems are likely to achieve.

Overall, the results of the experiment confirmed the expected benefits of uncertainty communication for human-automation interaction. As such, knowledge of system uncertainties allowed users to calibrate their trust in agreement with the true capabilities of the system. Changes in operator trust, as indicated by monitoring behaviour and questionnaire responses, thereby changed in agreement with the uncertainty level, with increasing uncertainties resulting in progressively decreasing trust. In agreement with related work (Bisantz, Marsiglio and Munch 2005), this indicates that users can distinguish several levels of uncertainties and utilise this knowledge to their benefit, i. e. successively increased engagement in NDRTs with decreasing uncertainty. Binary uncertainty communication, as utilised by Beller, Heesen and Vollrath (2013), would require either complete or no attention on the FRD, prompting the need for a carefully determined threshold. Such an implementation would function as an alarm that, depending on its frequency, may invoke the cry wolf effect (see Subsection 2.8.1) (Yang et al. 2017). In accordance with related work (Seppelt and Lee 2007), providing continuous information about the state of the automation circumvents this problem and allows for a more fluent interaction between user and system.

Besides the discussed benefits for human-automation interaction, the uncertainty display afflicted additional monitoring requirements on users, leading to increased workload and higher glance off-road times prior to critical situations. This prompted more research

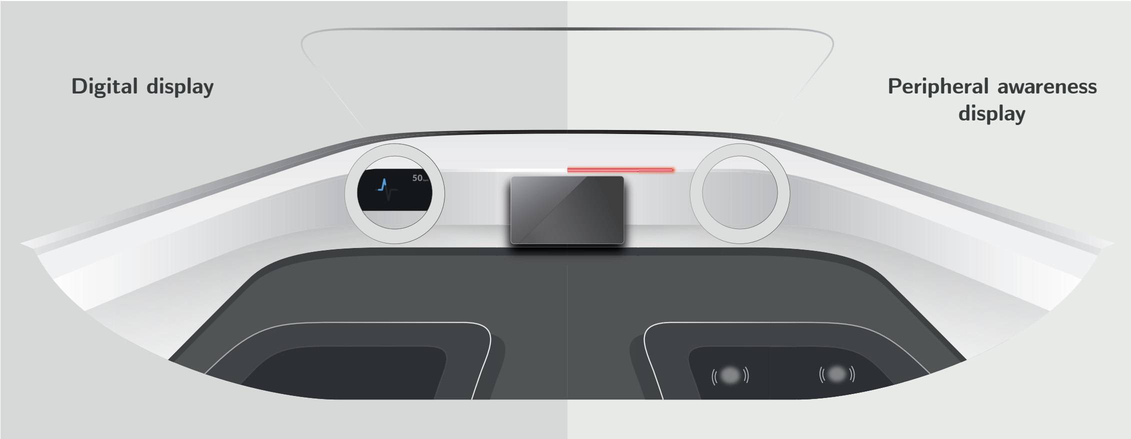


Figure 10.1: Uncertainty communication using the developed digital instrumental cluster display (left) and the peripheral awareness display (right)

efforts to develop an uncertainty display that mitigates the attentional requirements while maintaining the benefits for trust calibration.

10.2.2 Research Question 2: How can the systematic evaluation of peripheral awareness displays be ensured?

By definition, peripheral awareness displays allow users to perceive information of a secondary task while attending a primary task with focused attention. Consequently, such displays were proposed as a means for reducing the attentional demands of uncertainty displays. As a prerequisite of evaluating a peripheral awareness display for uncertainty communication in terms of its effects on the dependent variables investigated in RQ 1 (see Subsection 10.2.1), a suitable prototype display was first to be developed and evaluated.

Existing research has proposed a series of methods and frameworks for evaluating such displays, however, no publications known to the author have provided a systematic quantifiable evaluation framework and process.

Addressing this shortcoming, a systematic literature review was conducted to identify relevant criteria, dimensions, and methods for evaluating peripheral awareness displays. In a second step, the identified factors were combined with existing models of usability and user experience (see Section 2.3) to develop a comprehensive framework, the Peripheral Awareness Display Evaluation (PADE) framework (see Figure 5.5). Similar to the CUE model proposed by Thüring and Mahlke (2007), the PADE framework incorporates instrumental and non-instrumental qualities, emotional responses, and an overall appraisal. The weighting of the individual requirements is quantified based on design dimensions, specifically aesthetic emphasis and information criticality. This allows for a context-specific numeric evaluation of concept displays and an overall score pertaining to the user experience the displays afford. In practice, the evaluation is performed using a questionnaire, whereby its individual items were derived from related work (see Table 5.5) (Nielsen 1994b; Matthews, Rattenbury and Carter 2007; Shami, Lshed and Klein 2005; Matthews, Dey et al. 2004; ISO 2010; Minge 2018b). In combination with a heuristic evaluation, the developed PADE framework was subsequently implemented in a systematic development process based on the user-centred design process outlined in ISO 9241-210:2010 (ISO 2010) and the Double Diamond design process (Design Council 2018).

The developed process and framework serve as blueprints for researchers and designers aiming to develop and systematically evaluate peripheral awareness displays for various applications (see Section A.4). In contrast to existing frameworks and methods (Matthews, Rattenbury and Carter 2007; McCrickard, Chewar et al. 2003; Shami, Lshed and Klein 2005), the PADE framework allows for a quantified evaluation that facilitates the comparison of several display variants. Moreover, the developed process (see Figure 5.8) highlights how this framework can be implemented into a user-centred design process. Within this research project, the PADE framework and process served as underpinnings for subsequent research efforts that aimed to reduce the attentional demands of uncertainty displays.

10.2.3 Research Question 3: How can uncertainty information be intuitively conveyed using peripheral awareness displays?

Using the previously designed PADE process (see Subsection 10.2.2), RQ 3 was addressed in a series of laboratory-based experiments and evaluations.

As specified by the PADE framework, the usage context determines the specification of the design dimensions and thereby affects the weighting of the evaluation criteria. Consequently, a context of use analysis was performed at the start to establish guidelines for the development of a peripheral awareness display. With users of conditionally automated driving systems likely engaged in NDRTs that require focal visual and auditory resources, the vibro-tactile and peripheral visual channels were selected as modalities (see Section 6.1) to address unoccupied perceptual channels and thereby facilitate multitasking (see Subsection 2.4.2). The existing body of research lacked an investigation of variables for conveying uncertainties using these modalities, prompting the need for additional research.

Addressing these shortcomings, a set of visual and vibro-tactile variables were analysed in terms of their suitability for conveying different levels of uncertainties. The visual variables hue and size were found to be intuitive for conveying several levels of uncertainties. In agreement with related work, participants perceived animation-based variables as most interrupting (Mueller et al. 2014). Regarding vibro-tactile stimuli, particularly gradual increases in amplitude and rhythms consisting of long vibrations separated by short intervals were considered intuitive for conveying changes in uncertainty towards higher levels.

Based on these results, a first design iteration was developed that was subsequently evaluated in a heuristic evaluation (see Subsection 6.2.2). Using the developed PADE framework, the second iteration was then evaluated in a driving simulator study against the digital instrument cluster display employed to address RQ 1 (see Subsection 10.2.1). Thereby, the evaluation incorporated the same visual search task as the experiment outlined in Chapter 4. Both the aggregated UX scores and the scores of individual subscales indicated that the developed peripheral awareness display was preferred by users. In particular, the items relating to attentional requirements highlighted that the peripheral awareness display allowed users to be peripherally aware of uncertainty information while focusing on another task. Thus, the prerequisite for evaluating the implications of peripheral uncertainty communication on parameters of human-automation interaction was fulfilled.

10.2.4 Research Question 4: What are the implications of peripheral uncertainty communication on key parameters of human-automation interaction relative to the use of digital displays?

The experimental work conducted to address RQ 4 connects the findings of RQs 2 and 3 back to those of RQ 1. As such, a driving simulator study was conducted to assess how the developed peripheral awareness display compares with the initially deployed digital instrument cluster display in terms of key human-automation interaction aspects, in particular attention and workload.

While both display types supported trust calibration, eye tracking data suggests that the peripheral awareness display allowed users to perceive the conveyed information with fewer attentional requirements. Consequently, information access effort was reduced, manifesting itself in lower workload scores relative to the digital instrument cluster display. Confirming previous work (Seppelt, Seaman et al. 2017), decreased glance-off road times led to improved driving safety as indicated by increased MTTC values. Contrary to expectations, the reduced attentional demands of the uncertainty display did not translate to a significantly improved NDRT performance. For similar uncertainty levels, the NDRT performance was, however, better than that observed in the previous experiment (see Chapter 4) when no uncertainty display was present. This suggests that an increased practice as a consequence of the within-subjects design may have led to generally higher NDRT performance scores irrespective of the display type. Overall, the peripheral awareness display showed improvements in regard to workload and takeover performance metrics while allowing a similarly accurate judgement of the current uncertainty level as the digital instrument cluster display. In agreement with related work (Loecken, Heuten and Boll 2015), this highlights the benefits of conveying information by addressing previously unoccupied channels (Wickens 2008). Designers of human-machine systems are therefore encouraged to systematically consider the context of use when iterating display designs. This involves a consideration of the intended usage context during evaluation. As highlighted by the results of this experiment, the PADE framework and process provide a valid blueprint for guiding the development of peripheral awareness displays.

In practice, peripheral awareness displays for uncertainty communication should be combined with a driver monitoring system to determine appropriate salience levels. The display feedback should thereby be calibrated based on an assessment of the condition and activities of the driver.

10.2.5 Research Question 5: What are the qualitative and quantitative implications of increasing the functional detail of uncertainty information?

The experimental work conducted in connection with RQs 1–4 considered uncertainty as a system-wide metric and evaluated aspects pertaining to the communication method such as modality. Previous work has suggested that increases in the functional detail of information may lead to more appropriate trust (Lee and See 2004) and mitigate trust contagion effects (Rice and Geels 2010). To the knowledge of the author, no research efforts have investigated this in the context of driving automation.

Addressing this research gap, a driving simulator study was conducted whereby the functional detail of the uncertainty display was varied. Despite a lack of observable benefits to takeover performance, approximately half of participants valued the additional detail, stating that it aided system awareness and helped them to anticipate function-specific failures. In contrast, a similar number of participants preferred the reduced display with a minimal functional detail, considering the additional information unnecessary. An analysis of interview responses revealed that workload was a main factor when rejecting the increased functional detail of information. In agreement with the conclusions relating to RQ 1, display designs incorporating additional information should aim to prevent workload increments.

10.2.6 Research Question 6: What visualisation layouts and variables are best suited for conveying uncertainties with a higher functional detail?

The results of the experimental work conducted relating to RQ 5 suggested that an improved display design may invoke the anticipated benefits of increasing the functional detail of uncertainty information by reducing operator workload. Addressing this, an augmented reality (AR)-based uncertainty display layout was designed to lower the cognitive effort required to link the displayed information to the environment and reduce off-road glance time (Bubb et al. 2015). Moreover, a series of abstract visual variables were evaluated in terms of their guessability for conveying uncertainties in regard to different vehicle functions. In agreement with the experimental results recorded in connection with RQ 3, the variable hue was found to be most guessable for conveying uncertainties, as indicated by consistent sorting directions and high preference scores. Similarly, animation-based variables received high ratings, whereby pulses were preferred over more complex movements. As there were no main effects of the driving function for which the visual variables were implemented, it can be concluded that these variables work equally well in other implementations. As such, the results have implications for all displays showing content varying in uncertainty or urgency.

As the response time for the sorting task predicted preference ratings, the results confirm the notion that familiar cues such as colour hue facilitate intuitive cognition and thereby promote a faster information processing (see Subsection 2.4.1).

10.3 Contributions

This research project has resulted in a series of contributions to the field of human-automation interaction. Table 10.1 summarises these contributions with references to each chapter they pertain to.

Table 10.1: Summary of contributions derived from this research project

Contribution statement	**Chapter**
Consolidation of existing research into an integral model that illustrates how uncertainty communication affects situation awareness under consideration of related aspects such as workload or NDRTs	2
Uncertainty communication supports trust calibration, SA, and takeover performance	4, 7
Subjective trust indications are linked to the monitoring frequency of driving-related AOIs and NDRT task performance	4, 7
Using a digital instrument cluster display for conveying uncertainties increases workload and off-road glance time prior to critical events	4

Table 10.1: (Continuation) Summary of contributions derived from this research project

Contribution statement	Chapter
Usability, error visibility, interruption, noticeability, reaction, appeal, emotional responses, and satisfaction are key evaluation criteria for peripheral awareness displays	5
Aesthetic emphasis, information criticality, information capacity, representational fidelity, and personalisation are key design dimensions for categorising peripheral awareness displays	5
Specification of the PADE framework for the systematic development and evaluation of peripheral awareness displays	5
The variables hue (blue to red), size (small to large), and pulse (slow to fast) are well-suited for conveying uncertainties using lights	6
Animation-based variables should be reserved for higher levels of urgency	6
Vibro-tactile stimuli should only be used to convey increases in uncertainty, not decreases	6
The vibro-tactile variables amplitude and rhythm are well-suited for conveying increases in uncertainty	6
Using an uncertainty display that addresses the tactile and peripheral visual channel improves takeover performance while reducing workload relative to the use of a focal visual display	7
Workload is a key inhibitor for increasing the functional detail of uncertainty information	8
Preference regarding the degree of functional detail is highly individual	8
The visual variable hue and animation-based variables are well-suited for conveying information varying in urgency using an AR display	9

10.4 Limitations and Future Work

The results of this research project are constrained by a number of limitations that can serve as starting points for future work in the field.

The experiments presented in Chapters 4 and 7 involved the analysis of participant's eye movements to derive a measure for monitoring behaviour. The data was collected using a high-grade head-mounted eye tracking system, Tobii Pro Glasses 2 (see Subsection 3.2.3.2). With approximately 13,800 references in research publications indexed in Google Scholar (last update: 24 July 2019), Tobii is the most frequently mentioned manufacturer of eye tracking devices in research applications. Despite its popularity in the research community, the eye tracking glasses used in the context of this thesis were not without limitations. Tobii Pro Glasses 2 use the eye tracking technique *corneal reflection* in which a near-infrared light source illuminates the outer surface of the cornea and an image of the eye is captured that shows the light reflections in the pupil and on the cornea (Cognolato, Atzori and Müller 2018). Based on the geometrical features of these reflections, the position of fixation points can be calculated (Tobii AB 2018a). The proprietary eye tracking software Tobii Pro Lab (Tobii AB 2017) was used to map the identified fixation points to a photograph of the experimental scene. Due to differences in illumination between the projection screen and the vehicle cockpit ($<$300 lx) as well as dynamically changing contrasts on the projection screens, the automatic mapping feature provided results with limited confidence, prompting the manual review of mapped fixation points or – in some cases – the exclusion of participants if sufficient quality was not achieved. For instance, data from four participants were excluded from the eye tracking measures reported in Chapter 7. A more robust approach would have been to locally improve the illumination of the cockpit to reduce differences in lighting and, consequently, improve the confidence of the mapping feature. However, this would not have eliminated the issue of dynamically changing contrasts on the projection screen.

As discussed in Subsection 3.2.3.1, limitations concerning external validity arise when using driving simulators instead of real-world driving situations for data collection. In the context of this research project, the external validity of the driving simulator was particularly restricted by the limitations of STISIM Drive 3 (Systems Technology 2017). Differing from related work (Madigan, Louw and Institute 2018), driver-initiated takeovers could not be implemented. Instead, the system had to initiate the transfer in control. This contrasts driving automation systems that are currently available in commercial on-road vehicles. To deactivate systems like ACC, drivers can overrule the system by providing input to either brake or accelerator pedal. Knowledge of uncertainties may lead drivers to initiate a takeover themselves, thus preempting stressful system-initiated TORs (see Chapters 4 and 7). Future investigations should therefore grant participants the flexibility to initiate a takeover and assess if this mitigates the derogative effects of TORs on physiological measures such as heart rate or HRV. Moreover, partial takeovers could not be implemented using STISIM Drive 3. This would have been of interest for further investigations relating to the functional detail of uncertainty information (see Chapters 8 and 9). Besides, the required approximation of MTTC values (see Subsection 3.2.3.1) must be mentioned as a further limitation regarding the employed simulation software.

Another limitation pertaining to the use of a driving simulator is associated with the environmental indicator for uncertainties. Corresponding with related work (Louw, Markkula et al. 2017; Helldin et al. 2013), changes in fog density were used as a visible cause for increased system uncertainties. In contrast to real driving situations, fog was implemented as a two-dimensional grey layer that occluded the driving scene. It can be argued that in a three-dimensional scene, the effects would differ, thus limiting external validity. Moreover, commercially available automated driving systems can be expected to fail only on very rare occasions, perhaps once for several thousand kilometres (Favarò, Eurich and Nader 2018). Consequently, emergency TORs are going to be rare – and less frequent than in the experiments conducted in this research project. Considering that complacency and vigilance increase with time (Parasuraman and Manzey 2010), users cannot be expected to monitor an uncertainty display – or in fact any driving-related instrument – on a sustained basis (Large, Burnett, Morris and Muthumani 2017). While the peripheral awareness display developed as part of this project aimed to account for this, future investigations should assess how behaviour changes with prolonged faultless automation performance.

In this context, modalities other than those investigated in this project may prove useful to unobtrusively convey content over an extended period of time. As such, the kinaesthetic-vestibular (Cramer et al. 2017) or the olfactory channel (Wintersberger, Dmitrenko et al. 2019) could be addressed to communicate information about the system state.

In addition to the provision of uncertainty information, system transparency could further be increased by providing specific reasons for that uncertainty (Koo et al. 2015). This might enable partial system deactivations that require drivers to only takeover part of the DDT, e. g. the lateral component. The detail of conveyed information could also be dynamically adapted depending on the state of the driver (see Subsection 2.5.3.2). Thus, actively monitoring drivers might be interested in more detailed information, whereas users engaged in NDRTs likely benefit most from an overall metric. In reference to the interaction concept of cooperative guidance and control (see Subsection 2.5.3.1), awareness of system uncertainties may also allow users to judge when it is appropriate to grant more control to the system and when users should themselves have more influence.

Moreover, the results were obtained from single-visit experiments with limited exposure to the technologies and interfaces in question. Consequently, participants had little time to get accustomed to their new role as passive monitor of automated driving systems (Large, Burnett, Salanitri et al. 2019). Future research should employ longitudinal designs in order to assess the long-term effects of uncertainty communication.

Finally, future investigations should further elaborate on meta-uncertainties regarding the conveyed uncertainty information (Beller, Heesen and Vollrath 2013). The peripheral awareness display employed in Chapter 7 communicated meta-uncertainties using a binary indicator for the working state of the uncertainty display. More research is required to investigate how systems can determine and convey the confidence of uncertainty values.

10.5 Overall Conclusion

The concurrent theme throughout all investigations of this research project was the significance of considering the essential HF aspects human, machine, and environment in conjunction (see Figure 2.9). Well-trained individuals such as pilots may be expected to remain somewhat involved in the flying task through regular checks (Trösterer, Meschtscherjakov et al. 2017). The same should not be expected of drivers who were shown to readily engage in NDRTs, even without automated driving systems available (see Subsection 2.5.2). Consequently, machine displays should be designed with the usage context in mind.

As such, the benefits of uncertainty communication for human-automation interaction can be maximised through a context-depending handling of attentional resources. Specifically, display designs should aim to reduce workload by addressing previously unoccupied perceptual channels and by using intuitively processed signifiers.

So long as faultlessly performing automated driving systems are not available, automation will not replace human performance, but rather change it. Designers of human-automation interfaces should strive to make this change as user-centred as possible.

References

AAA Foundation for Traffic Safety (2018). *Traffic Safety Culture Index.* URL: https://publicaffairsresources.aaa.biz/wp-content/uploads/2018/03/AAAFTS-TSCI-Fact-Sheet.pdf (visited on 23/02/2019).

Air Transport Association of America (1989). *ATA Human Task Force report.* Tech. rep.

Aircraft Accident Investigation Commission (Ministry of Transport Japan) (1996). *Aircraft Accident Investigation Report. China Airlines Flight 140. Airbus Industrie A300B4-622R, B1816. Nagoya Airport. April 26, 1994.*

Alkhafaj, Alaa S. A. and Sanaz Fallahkhair (2014). 'Smart Ambient: Development of Mobile Location Based System to Support Informal Learning in the Cultural Heritage Domain'. In: *2014 IEEE 14th International Conference on Advanced Learning Technologies.* Athens, Greece: IEEE, pp. 774–776. ISBN: 9781479940387. DOI: 10.1109/ICALT.2014.227.

Altendorf, Eugen et al. (2016). 'Joint Decision Making and Cooperative Driver-Vehicle Interaction during Critical Driving Situations'. In: *I-Com* 15.3, pp. 265–281. ISSN: 2196-6826. DOI: 10.1515/icom-2016-0033.

Andre, Anthony D. and Henry A. Cutler (1998). 'Displaying Uncertainty in Advanced Navigation Systems'. In: *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, pp. 31–35.

Angelucci, Daniela, Stefania Di Paolo and Laura Tarantino (2009). 'Designing a glanceable peripheral display for severity-based multiple alarm notification'. In: *Proceedings - 2009 2nd Conference on Human System Interactions, HSI '09*, pp. 242–246. DOI: 10.1109/HSI.2009.5090985.

Aparicio, Manuela and Carlos J. Costa (2014). 'Data Visualization'. In: *Communication Design Quarterly* 3.1, pp. 7–11.

Aral Aktiengesellschaft (2017). *Aral Studie: Trends beim Autokauf 2017.* URL: https://www.aral.de/content/dam/aral/Presse%20Assets/pdfs-Broschueren/Aral-Studie-Trends-beim-Autokauf-2017.pdf (visited on 24/02/2019).

Aranyi, Gabor and Paul van Schaik (2016). 'Testing a Model of User-Experience with News Websites'. In: *Communications in Information Literacy* 67.7, pp. 1555–1575. ISSN: 19335954. DOI: 10.1002/asi. arXiv: 0803.1716.

Ariely, Dan (2001). 'Seeing sets: representation by statistical properties'. In: *Psychological Science* 12.2, pp. 157–162. ISSN: 0956-7976. DOI: 10.1111/1467-9280.00327.

Ayres, Lioness (2008). 'Semi-Structured Interview'. In: *The SAGE Encyclopedia of Qualitative Research Methods.* Ed. by Lisa M. Given. Thousand Oaks: SAGE Publications, Inc., pp. 811–812. DOI: http://dx.doi.org/10.4135/9781412963909.n420.

Bagheri, Nasrine and Greg A. Jamieson (2004). 'The Impact of Context-related Reliability on Automation Failure Detection and Scanning Behaviour'. In: *2004 IEEE International Conference on Systems, Man and Cybernetics*. The Hague, Netherlands: IEEE, pp. 212–217. ISBN: 0780385667. DOI: 10.1109/ICSMC.2004.1398299.

Bailey, Nathan R. and Mark W. Scerbo (2007). 'Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust'. In: *Theoretical Issues in Ergonomics Science* 8.4, pp. 321–348. ISSN: 1463-922X. DOI: 10.1080/14639220500535301.

Bainbridge, Lisanne (1983). 'Ironies of automation'. In: *Automatica* 19.6, pp. 775–779. ISSN: 0005-1098. DOI: 10.1016/0005-1098(83)90046-8.

Balfe, Nora, Sarah Sharples and John R. Wilson (2018). 'Understanding Is Key: An Analysis of Factors Pertaining to Trust in a Real-World Automation System'. In: *Human Factors* 60.4, pp. 477–495. ISSN: 15478181. DOI: 10.1177/0018720818761256.

Banks, Victoria A. et al. (2018). 'Is partially automated driving a bad idea? Observations from an on-road study'. In: *Applied Ergonomics* 68.November 2017, pp. 138–145. ISSN: 18729126. DOI: 10.1016/j.apergo.2017.11.010.

Bao, Yan et al. (2018). 'Attentional modulation of speed-change perception in the perifoveal and near-peripheral visual field'. In: *PLOS ONE* August, pp. 1–17. DOI: https://doi.org/10.1371/journal.pone.0203024.

Barker, Jon (2017). *Introduction to Deep Learning on GPUs*. URL: https://nvidia.qwiklab.com/tags/Deep%20Learning (visited on 01/03/2017).

Bates, Douglas et al. (2018). *Linear Mixed-Effect Models using 'Eigen' and S4*. New York, NY, US: Version 1.1-17. URL: https://cran.r-project.org/web/packages/lme4/lme4.pdf.

Battaglia, Contributors Mike (2008). 'Convenience Sampling'. In: *Encyclopedia of Survey Research Methods*. Ed. by Paul J. Lavrakas. Thousand Oaks: SAGE Publications, Inc., p. 149. ISBN: 9781412963947. DOI: http://dx.doi.org/10.4135/9781412963947.n105.

Bazilinskyy, Pavlo et al. (2018). 'Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays'. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 56, pp. 82–98. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.13. URL: https://doi.org/10.1016/j.trf.2018.04.001.

BBC News (2009). *Man follows sat nav to cliff edge*. URL: http://news.bbc.co.uk/2/hi/uk_news/england/bradford/7962212.stm (visited on 27/02/2016).

BEA (2012). 'Final Report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro - Paris'. In: *Safety Investigations*, p. 223. ISSN: 0029-6570. DOI: 10.2172/875800.

Bellem, Hanna et al. (2018). 'Comfort in automated driving: An analysis of preferences for different automated driving styles and their dependence on personality traits'. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 55, pp. 90–100. ISSN: 13698478. DOI: 10.1016/j.trf.2018.02.036.

Beller, Johannes, Matthias Heesen and Mark Vollrath (2013). 'Improving the Driver Automation Interaction: An Approach Using Automation Uncertainty'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 55.6, pp. 1130–1141. ISSN: 0018-7208. DOI: 10.1177/0018720813482327.

Benjamini, Yoav and Yosef Hochberg (1995). 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing'. In: *Journal of the Royal Statistical Society Series B* 57.1, pp. 289–300.

Bertin, Jacques (1967). *Semiology of Graphics: Diagrams, Networks, Maps.* University of Wisconsin.

Bertoncello, Michele and Dominik Wee (2015). *Ten ways autonomous driving could redefine the automotive world.* URL: http://www.mckinsey.com/industries/automotive-and-assembly/our-insights/ten-ways-autonomous-driving-could-redefine-the-automotive-world (visited on 20/02/2017).

Bevan, Nigel and Miles Macleod (1994). 'Usability measurement in context'. In: *Behaviour & Information Technology* 13.1-2, pp. 132–145. ISSN: 0144-929X. DOI: 10.1080/01449299408914592.

Billings, Charles E. (1997). *Aviation Automation: The Search for a Human-Centered Approach.* Mahwah, N.J.: Lawrence Erlbaum Associates, p. 355. ISBN: 0805821260.

BIOPAC Systems Inc. (2013). *Electrode Lead Set SS2L.* URL: https://www.biopac.com/wp-content/uploads/SS2L.pdf (visited on 20/03/2019).

– (2017). *MP35 Four Channel Data Acquisition System.* URL: https://www.biopac.com/wp-content/uploads/MP35.pdf (visited on 20/03/2019).

– (2018a). *Pulse Photoplethysmogram Transducer SS4LA.* URL: https://www.biopac.com/wp-content/uploads/Pulse-Transducer.pdf (visited on 20/03/2019).

– (2018b). *Respiratory Effort Transducer SS5LB.* URL: https://www.biopac.com/wp-content/uploads/SS5LB.pdf (visited on 20/03/2019).

Bisantz, Ann M., Stephanie Schinzing Marsiglio and Jessica Munch (2005). 'Displaying Uncertainty: Investigating the Effects of Display Format and Specificity'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 47.4, pp. 777–796. ISSN: 1547-8181. DOI: 10.1518/001872005775570916.

Bland, Martin (2015). *An Introduction to Medical Statistics.* Oxford University Press, p. 448. ISBN: 9780199589920.

Boelhouwer, A. et al. (2019). 'Should I take over? Does system knowledge help drivers in making take-over decisions while driving a partially automated car?' In: *Transportation Research Part F: Traffic Psychology and Behaviour* 60, pp. 669–684. ISSN: 13698478. DOI: 10.1016/j.trf.2018.11.016.

Borojeni, Shadan S., Abdallah El Ali et al. (2016). 'Peripheral Light Cues for In-Vehicle Task Resumption'. In: *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, 67:1–67:4. DOI: 10.1145/2971485.2971498.

Borojeni, Shadan S., Lewis Chuang et al. (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*. October. Ann Arbor, MI, USA: ACM, pp. 237–244. ISBN: 9781450345330. DOI: http://dx.doi.org/10.1145/3003715.3005409.

Bratzel, Stefan (2018). *Der Abgasskandal und die Vertrauenskrise im Automobilmarkt – Ursachen, Lösungen und Auswirkungen auf den Verbraucher*. URL: https://www.vzbv.de/sites/default/files/downloads/2018/09/13/cam_gutachten_vertrauenskrise_im_automobilmarkt_v1.1_0.pdf (visited on 29/03/2019).

Breznitz, Shlomo (1984). *Cry Wolf: The Psychology of False Alarms*. London: Lawrence Erlbaum Associates. ISBN: 9780898592962.

Brinkmann, Svend (2008). 'Interviewing'. In: *The SAGE Encyclopedia of Qualitative Research Methods*. Ed. by Lisa M. Given. Thousand Oaks: SAGE Publications, Inc., pp. 471–472. DOI: http://dx.doi.org/10.4135/9781412963909.n239.

Bryman, Alan and Emma Bell (2015). *Business Research Methods*. 4th editio. Oxford: Oxford University Press, p. 816. ISBN: 0199668647.

Brysbaert, Marc and Michaël Stevens (2018). 'Power Analysis and Effect Size in Mixed Effects Models: A Tutorial'. In: *Journal of Cognition* 1.1, pp. 1–20. ISSN: 2514-4820. DOI: 10.5334/joc.10.

Bubb, Heiner et al. (2015). *Automobilergonomie*. Wiesbaden: Springer Vieweg, pp. 663–684. ISBN: 9783834818904. DOI: 10.1007/978-3-642-13505-7.

Carrasco, Marisa (2011). 'Visual attention: The past 25 years'. In: *Vision Research* 51.13, pp. 1484–1525. ISSN: 00426989. DOI: 10.1016/j.visres.2011.04.012. arXiv: 156869.

Carroll, John M and Judith Reitman Olson, eds. (1987). *Mental Models in Human-computer Interaction: Research Issues About What the User of Software Knows*. Washington, DC, USA: National Academy Press.

Carsten, Oliver et al. (2012). 'Control Task Substitution in Semiautomated Driving: Does It Matter What Aspects Are Automated?' In: *Human Factors: The Journal of the Human*

Factors and Ergonomics Society 54.5, pp. 747–761. ISSN: 0018-7208. DOI: 10.1177/0018720812460246.

Casner, Stephen M., Richard W. Geven et al. (2014). 'The Retention of Manual Flying Skills in the Automated Cockpit'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 56.8, pp. 1506–1516. ISSN: 0018-7208. DOI: 10.1177/0018720814535628.

Casner, Stephen M., Edwin L. Hutchins and Donald A. Norman (2016). 'The Challenges of Partially Automated Driving'. In: *Communications of the ACM* 59.5, pp. 70–77. ISSN: 00010782. DOI: 10.1145/2830565.

Casner, Stephen M. and Jonathan W. Schooler (2014). 'Thoughts in Flight: Automation Use and Pilots' Task-Related and Task-Unrelated Thought'. In: *Human Factors* 56.3, pp. 433–442. DOI: 10.1177/0018720813501550.

Cellan-Jones, Rory (2018). *Car insurers warn on 'autonomous' vehicles*. URL: https://www.bbc.co.uk/news/technology-44439523 (visited on 12/06/2018).

Chancey, Eric T. et al. (2017). 'Trust and the Compliance-Reliance Paradigm: The Effects of Risk, Error Bias, and Reliability on Trust and Dependence'. In: *Human Factors* 59.3, pp. 333–345. ISSN: 15478181. DOI: 10.1177/0018720816682648.

Chandler, Daniel (2017). *Semiotics: The Basics*. Third edit. London: Routledge, Taylor & Francis Group, p. 331. ISBN: 9781138232938.

Chang, Wonsuk, Wonil Hwang and Yong Gu Ji (2011). 'Haptic seat interfaces for driver information and warning systems.' In: *International Journal of Human-Computer Interaction* 27.12, pp. 1119–1132. ISSN: 1044-7318. DOI: 10.1080/10447318.2011.555321.

Charness, Gary, Uri Gneezy and Michael A. Kuhn (2012). 'Experimental methods: Between-subject and within-subject design'. In: *Journal of Economic Behavior and Organization* 81.1, pp. 1–8. ISSN: 01672681. DOI: 10.1016/j.jebo.2011.08.009.

Chavaillaz, Alain, David Wastell and Jürgen Sauer (2016). 'System reliability, performance and trust in adaptable automation'. In: *Applied Ergonomics* 52, pp. 333–342. ISSN: 0003-6870. DOI: 10.1016/j.apergo.2015.07.012.

Chen, Jessie Y. C., Katelyn Procci et al. (2014). 'Situation Awareness – Based Agent Transparency'. In: April, pp. 1–29.

Chen, Jessie Y.C., Shan G. Lakhmani et al. (2018). 'Situation awareness-based agent transparency and human-autonomy teaming effectiveness'. In: *Theoretical Issues in Ergonomics Science* 19.3, pp. 259–282. ISSN: 1464536X. DOI: 10.1080/1463922X.2017.1315750.

Chong, Sang Chul and Anne Treisman (2003). 'Representation of statistical properties'. In: *Vision Research* 43.4, pp. 393–404. ISSN: 00426989. DOI: 10.1016/S0042-6989(02)00596-5.

Chong, Sang Chul and Anne Treisman (2005). 'Attentional spread in the statistical processing of visual displays'. In: *Perception & psychophysics* 67.1, pp. 1–13. ISSN: 0031-5117. DOI: 10.3758/BF03195009.

Clamann, Michael and Mary L. Cummings (2016). 'What is the automation doing? Mode awareness problems catch Tesla by surprise'. In: *Ergonomics in Design: The Quarterly of Human Factors Applications.*

Cognolato, Matteo, Manfredo Atzori and Henning Müller (2018). 'Head-mounted eye gaze tracking devices: An overview of modern devices and recent advances'. In: *Journal of Rehabilitation and Assistive Technologies Engineering* 5. ISSN: 2055-6683. DOI: 10.1177/2055668318773991.

Cohen, Marvin S., Raja Parasuraman and Jared T. Freeman (1997). *Trust in Decision Aids: A Model and a Training Strategy.* Arlington, VA: Cognitive Technologies, Inc. for US Army Aviation and Troop Command. DOI: 10.1.1.90.2591.

Connor, Charles E., Howard E. Egeth and Steven Yantis (2004). 'Visual attention: Bottom-up versus top-down'. In: *Current Biology* 14.19, pp. 850–852. ISSN: 09609822. DOI: 10.1016/j.cub.2004.09.041.

Consumer Reports (2015). *Latest Tesla Model S Software Update Includes Autopilot and a Taste of Autonomous Driving.* URL: https://www.consumerreports.org/cars-tesla-model-s-software-update-7-autopilot-autonomous-driving/ (visited on 12/12/2018).

Cooley, Mike (1987). 'Human centred Systems: An Urgent Problem for Systems Designers'. In: *AI & Society* 1, pp. 37–46.

Copeland, Michael (2016). *What's the Difference Between Artificial Intelligence, Machine Learning, and Deep Learning?* URL: https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/ (visited on 22/02/2017).

Corbetta, Maurizio and Gordon L. Shulman (2002). 'Control of goal-directed and stimulus-driven attention in the brain'. In: *Nature Reviews Neuroscience* 3.3, pp. 201–215. ISSN: 14710048. DOI: 10.1038/nrn755.

Cramer, Stephanie et al. (2017). 'Design Possibilities for Vehicle Roll Motions as Feedback for the Driver during Automated Driving'. In: *AutomotiveUI '17 Adjunct.* Oldenburg, Germany: ACM, pp. 152–157. DOI: 10.1145/3131726.3131746.

Creaser, Janet and Gregory M. Fitch (2015). 'Human Factors Considerations for the Design of Level 2 and Level 3 Automated Vehicles'. In: *Road Vehicle Automation 2.* Springer. DOI: 10.1007/978-3-319-19078-5_8.

Cummings, Mary L. and Jason Ryan (2013). *Who is in Charge? Promises and Pitfalls of Driverless Cars.* URL: http://hal.pratt.duke.edu/sites/hal.pratt.duke.edu/files/u7/TR%20news%20Cummings%20MAR14.pdf (visited on 29/11/2016).

Cuthbertson, Anthony (2019). *Tesla Driver Filmed 'Sleeping at the Wheel' of Semi-Autonomous Model X.* URL: https://www.independent.co.uk/life-style/gadgets-and-tech/news/tesla-driver-sleeping-video-autonomous-car-model-x-autopilot-a8756261.html (visited on 31/03/2019).

Damböck, Daniel et al. (2012). 'Evaluation of a Contact Analog Head-Up Display for Highly Automated Driving'. In: *4th International Conference on Applied Human Factors and Ergonomics.* San Francisco, CA, pp. 6011–6020.

Davenport, Mark and Don Harris (1992). 'Pilot Interaction With Cockpit Automation: Operational Experiences With the Flight Management System'. In: *The International Journal of Aviation Psychology* 2.4, pp. 271–280. ISSN: 1532-7108. DOI: 10.1207/s15327108ijap0204_5.

Davies, Alex (2018). *A Sleeping Tesla Driver Highlights Autopilot's Biggest Flaw.* URL: https://www.wired.com/story/tesla-sleeping-driver-dui-arrest-autopilot/ (visited on 12/03/2019).

Davies, David R. and Raja Parasuraman (1982). *The Psychology of Vigilance.* Academic Press, p. 300. ISBN: 978-0122061806.

Davis, Kadian et al. (2017). 'Effects of Ambient Lighting Displays on Peripheral Activity Awareness'. In: *IEEE Access* 5, pp. 9318–9335. ISSN: 21693536. DOI: 10.1109/ACCESS.2017.2703866.

Davson, Hugh and Edward S. Perkins (2018). 'Human eye'. In: *Encyclopaedia Britannica.* Ed. by Adam Augusty et al.

DeBolt, Daniel (2011). *Google's self-driving car in five-car crash.* URL: http://www.paloaltoonline.com/news/2011/08/08/googles-self-driving-car-in-five-car-crash (visited on 02/12/2016).

Department of Defense (2012). 'The Role of Autonomy in DoD Systems'. In: *DoD Defense Science Board* July, p. 125. ISSN: 01406736. DOI: 10.1016/S0140-6736(02)11924-6.

Design Council (2018). *The Design Process: What is the Double Diamond?* URL: https://www.designcouncil.org.uk/news-opinion/design-process-what-double-diamond (visited on 08/06/2018).

Diekfuss, Jed A., Paul Ward and Louisa D. Raisbeck (2017). 'Attention, workload, and performance: A dual-task simulated shooting study'. In: *International Journal of Sport and Exercise Psychology* 15.4, pp. 423–437. ISSN: 1557251X. DOI: 10.1080/1612197X.2015.1121508.

Dillon, Andrew (2001). 'Beyond Usability: Process, Outcome, and Affect in Human Computer Interactions.' In: *Canadian Journal of Information & Library Sciences* 26.4, pp. 57–70. ISSN: 1195096X.

Dogan, Ebru, Linda Steg and Patricia Delhomme (2011). 'The influence of multiple goals on driving behavior: The case of safety, time saving, and fuel saving'. In: *Accident Analysis and Prevention* 43.5, pp. 1635–1643. ISSN: 00014575. DOI: 10.1016/j.aap.2011.03.002.

Donges, Edmund (1982). 'Aspekte der aktiven Sicherheit bei der Fuehrung von Personenkraftwagen'. In: *Automobilindustrie* 27.2.

Drecki, Igor (2002). 'Visualization of Uncertainty in Geographical Data'. In: *Spatial Data Quality*. Ed. by W. Shi, P. Fisher and M. Goodchild. London: Taylor & Francis, pp. 140–159.

Dunning, David et al. (2003). 'Why people fail to recognise their own incompetence.' In: *Current Directions in Psychological Science: A Journal of the American Psychological Society* 12, p. 83. ISSN: 09637214. DOI: 10.1111/1467-8721.01235.

Dzindolet, Mary T. et al. (2003). 'The role of trust in automation reliance'. In: *International Journal of Human Computer Studies* 58.6, pp. 697–718. ISSN: 10715819. DOI: 10.1016/S1071-5819(03)00038-7.

Economist (2018). *Self-driving cars will profoundly change the way people live.* URL: https://www.economist.com/special-report/2018/03/01/self-driving-cars-will-profoundly-change-the-way-people-live?fsrc=scn/fb/te/bl/ed/selfdrivingcarswillprofoundlychangethewaypeopleliveadifferentworld (visited on 01/06/2018).

Endsley, Mica R. (1988a). 'Design and Evaluation for Situation Awareness Enhancement'. In: *Proceedings of the Human Factors Society 32nd Annual Meeting*. January. Santa Monica, CA: Human Factors and Ergonomics Society, pp. 97–101. DOI: 10.1177/154193128803200221.

– (1988b). 'Situation awareness global assessment technique (SAGAT)'. In: *Proceedings of the National Aerospace and Electronics Conference (NAECON*. Hawthorne, CA: IEEE, pp. 789–795. DOI: 10.1109/NAECON.1988.195097.

– (1995a). 'Measurement of Situation Awareness in Dynamic Systems'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37.1, pp. 65–84. ISSN: 0018-7208. DOI: 10.1518/001872095779049499.

– (1995b). 'Toward a Theory of Situation Awareness in Dynamic Systems'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37.1, pp. 32–64. ISSN: 00187208. DOI: 10.1518/001872095779049543.

– (2015). ‘Situation Awareness Misconceptions and Misunderstandings’. In: *Journal of Cognitive Engineering and Decision Making* 9.1, pp. 4–32. ISSN: 1555-3434. DOI: 10.1177/1555343415572631.

– (2017). ‘From Here to Autonomy: Lessons Learned From Human-Automation Research’. In: *Human Factors The Journal of the Human Factors and Ergonomics Society* 59.1, pp. 5–27. ISSN: 0018-7208. DOI: 10.1177/0018720816681350.

Endsley, Mica R. and D. J. Garland (2000). ‘Direct measurement of situation awareness: validity and use of SAGAT Development of Queries’. In: *Situation Awareness Analysis and Measurement*. Mahwah, N.J.: Lawrence Erlbaum Associates, pp. 1–21. ISBN: 0805821333.

Endsley, Mica R. and Debra G. Jones (2012). *Designing for Situation Awareness*. Boca Raton, FL, US: CRC Press, p. 393. ISBN: 9781420063585.

Endsley, Mica R. and Esin O. Kiris (1995). ‘The Out-of-the-Loop Performance Problem and Level of Control in Automation’. In: *Human Factors and Ergonomics Society* 37.2, pp. 381–394. ISSN: 00187208. DOI: 10.1518/001872095779064555.

Epilepsy Society (2018). *Photosensitive Epilepsy*. URL: https://www.epilepsysociety.org.uk/photosensitive-epilepsy (visited on 08/06/2018).

Erie Insurance (2017). *National survey from Erie Insurance finds majority think self-driving cars will eliminate distracted driving*. URL: https://www.erieinsurance.com/news-room/press-releases/2017/self-driving-distracted (visited on 23/02/2019).

Eriksson, Alexander, Victoria A. Banks and Neville A. Stanton (2017). ‘Transition to Manual: comparing simulator with on-road control transitions Eriksson’. In: *Accident Analysis and Prevention* 102, pp. 227–234. ISSN: 00014575. DOI: 10.1016/j.aap.2017.03.011.

Eriksson, Alexander and Neville A. Stanton (2017). ‘Takeover Time in Highly Automated Vehicles: Noncritical Transitions to and from Manual Control’. In: *Human Factors* 59.4, pp. 689–705. ISSN: 15478181. DOI: 10.1177/0018720816685832.

Erlingsson, Christen and Petra Brysiewicz (2017). ‘A hands-on guide to doing content analysis’. In: *African Journal of Emergency Medicine* 7.3, pp. 93–99. ISSN: 2211419X. DOI: 10.1016/j.afjem.2017.08.001.

Erp, Jan B. F. van (2002). ‘Guidelines for the Use of Vibro-Tactile Displays in Human Computer Interaction’. In: *Proceedings of Eurohaptics*, pp. 18–22. ISSN: 1097-6752. DOI: 10.1016/j.ajodo.2008.04.017.

European Road Transport Research Advisory Council (2015). *Automated Driving Roadmap*. URL: http://www.ertrac.org/uploads/documentsearch/id38/ERTRAC_Automated-Driving-2015.pdf (visited on 16/01/2017).

FAA (2013). *SAFO 13002, Manual Flight Operations*. Tech. rep. 13002. Washington, DC: U.S. Department of Transportation. URL: https://www.faa.gov/other_visit/

aviation_industry/airline_operators/airline_safety/safo/all_safos/media/2013/SAFO13002.pdf.

Faltalous, Sarah et al. (2018). 'Design Guidelines for Reliability Communication in Autonomous Vehicles'. In: *AutomotiveUI'18 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* September. DOI: 10.1145/3239060.3239072.

Favarò, Francesca, Sky Eurich and Nazanin Nader (2018). 'Autonomous vehicles' disengagements: Trends, triggers, and regulatory limitations'. In: *Accident Analysis and Prevention* 110, pp. 136–148. ISSN: 00014575. DOI: 10.1016/j.aap.2017.11.001.

Feys, Jos (2016). 'Nonparametric Tests for the Interaction in Two-way Factorial Designs Using R'. In: *The R Journal* 8.1, pp. 367–378. ISSN: 20734859. URL: https://journal.r-project.org/archive/2016/RJ-2016-027/RJ-2016-027.pdf.

Finger, Richard and Ann M. Bisantz (2002). 'Utilizing graphical formats to convey uncertainty in a decision-making task'. In: *Theoretical Issues in Ergonomics Science* 2.1, pp. 1–25. ISSN: 1463-922X. DOI: 10.1080/14639220110110324.

Fitch, Gregory M. et al. (2013). *The Impact of Hand-Held And Hands-Free Cell Phone Use on Driving Performance and Safety-Critical Event Risk (DOT HS811 757)*. Tech. rep. April. NHTSA, pp. 1–273. URL: https://www.distraction.gov/downloads/pdfs/the-impact-of-hand-held-and-hands-free-cell-phone-use-on-driving-performance-and-safety-critical-event-risk.pdf.

Flemisch, Frank, Klaus Bengler et al. (2014). 'Towards cooperative guidance and control of highly automated vehicles: H-Mode and Conduct-by-Wire.' In: *Ergonomics* 57.3, pp. 343–60. ISSN: 1366-5847. DOI: 10.1080/00140139.2013.869355.

Flemisch, Frank, Matthias Heesen et al. (2012). 'Towards a dynamic balance between humans and automation: Authority, ability, responsibility and control in shared and cooperative control situations'. In: *Cognition, Technology and Work* 14.1, pp. 3–18. ISSN: 14355558. DOI: 10.1007/s10111-011-0191-6.

Flemisch, Frank, Johann Kelsch et al. (2008). 'Cooperative Control and Active Interfaces for Vehicle Assistance and Automation'. In: *FISITA World Automotive Congress* 2, pp. 301–310.

Fortmann, Florian et al. (2015). 'Expert-based design and evaluation of an ambient light display to improve monitoring performance during multi-UAV supervisory control'. In: *2015 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision, CogSIMA 2015*, pp. 28–34. DOI: 10.1109/COGSIMA.2015.7107971.

Free, John (1977). 'Microprocessors: Tiny overseers than run everything – from cars to ovens'. In: *Popular Science* March, pp. 90–93.

Fridman, Lex (2017). 'Deep Learning for Self-Driving Cars'. In: *Lecture Notes*. Cambridge, Mass.: Massachussetts Institute of Technology. URL: http://selfdrivingcars.mit.edu/.

Gabbard, Joseph L., Gregory M. Fitch and Hyungil Kim (2014). 'Behind the glass: Driver challenges and opportunities for AR automotive applications'. In: *Proceedings of the IEEE* 102.2, pp. 124–136. ISSN: 00189219. DOI: 10.1109/JPROC.2013.2294642.

Gal, Yarin (2016). 'Uncertainty in Deep Learning'. PhD thesis. University of Cambridge.

Gartner, Inc. (2016). *Gartner's 2016 Hype Cycle for Emerging Technologies Identifies Three Key Trends That Organizations Must Track to Gain Competitive Advantage*. URL: http://www.gartner.com/newsroom/id/3412017 (visited on 27/03/2017).

Gasser, Tom M. and Daniel Westhoff (2012). *BASt-study: Definitions of Automation and Legal Issues in Germany*.

Geels-Blair, Kasha, Stephen Rice and Jeremy Schwark (2013). 'Using System-Wide Trust Theory to Reveal the Contagion Effects of Automation False Alarms and Misses on Compliance and Reliance in a Simulated Aviation Task'. In: *The International Journal of Aviation Psychology* 23.3, pp. 245–266. ISSN: 1050-8414. DOI: 10.1080/10508414.2013.799355.

Geiselman, Eric E., Christopher M. Johnson and David R. Buck (2013). 'Flight deck automation: Invaluable collaborator or insidious enabler?' In: *Ergonomics in Design: The Quarterly of Human Factors Applications* 21.July, pp. 22–26. ISSN: 1064-8046. DOI: 10.1177/1064804613491268.

Ginwalla, Neville Z. et al. (2012). 'Measurement of Trust in Humans in Hybrid Inspection for Different Levels of Error Randomness'. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 46.23, pp. 1920–1924. ISSN: 1541-9312. DOI: 10.1177/154193120204602311.

GISGeography (2017). *What is Geographic Information Systems (GIS)?* URL: http://gisgeography.com/what-gis-geographic-information-systems/ (visited on 31/05/2017).

GMC (2018). *Seat pulses alert you to potential hazards*. URL: https://my.gmc.com/learnAbout/safety-alert-seat/2018/Acadia (visited on 15/03/2019).

Gold, Christian, Daniel Damböck et al. (2013). '"Take over!" How long does it take to get the driver back into the loop?' In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 57.1, pp. 1938–1942. ISSN: 1541-9312. DOI: 10.1177/1541931213571433.

Gold, Christian, Moritz Körber et al. (2016). 'Taking Over Control From Highly Automated Vehicles in Complex Traffic Situations: The Role of Traffic Density'. In: *Human Factors:*

The Journal of the Human Factors and Ergonomics Society 58.4. DOI: 10.1177/0018720816634226.

Golightly, David et al. (2010). 'The role of situation awareness for understanding signalling and control in rail operations'. In: *Theoretical Issues in Ergonomics Science* 11.1-2, pp. 84–98. ISSN: 1463922X. DOI: 10.1080/14639220903009961.

Gollan, Jackie K. et al. (2010). 'How do depressed and healthy adults interpret nuanced facial expressions?' In: *Journal of Abnormal Psychology* 119.4, pp. 804–810. ISSN: 19391846. DOI: 10.1037/a0020234.

Gomes, Lee (2016). 'When Will Google's Self-Driving Car Really Be Ready? Three Super-Cool Technologies'. In: *IEEE Spectrum* 53.5, pp. 13–14.

Green, Paul (2013). 'Standard Definitions for Driving Measures and Statistics: Overview and Status of Recommended Practice J2944'. In: *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '13*. Eindhoven, Netherlands: ACM, pp. 184–191. DOI: http://dx.doi.org/10.1145/2516540.25165422.

Greenlee, Eric T., Patricia R. DeLucia and David C. Newton (2018). 'Driver Vigilance in Automated Vehicles: Hazard Detection Failures Are a Matter of Time'. In: *Human Factors* 60.4, pp. 465–476. ISSN: 15478181. DOI: 10.1177/0018720818761711.

Grubb, P. L. et al. (1994). 'Cognitive failure and perceived workload in vigilance performance'. In: *Human performance in automated systems: Current research and trends.* Ed. by Mustapha Mouloua and Raja Parasuraman. Hillsdale, NJ: Lawrence Erlbaum, pp. 115–121.

Guest, Greg, Arwen Bunce and Laura Johnson (2006). 'How Many Interviews Are Enough? An Experiment with Data Saturation and Variability'. In: *Field Methods* 18.1, pp. 59–82. ISSN: 1525822X. DOI: 10.1177/1525822X05279903.

Haeuslschmid, Renate et al. (2016). 'First Steps towards a View Management Concept for large-sized Head-Up Displays with Continuous Depth'. In: *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* Ann Arbor, MI, USA: ACM, pp. 1–8. ISBN: 9781450345330.

Halik, Łukasz (2012). 'The analysis of visual variables for use in the cartographic design of point symbols for mobile Augmented Reality applications'. In: *Geodesy and Cartography* 61.1, pp. 19–30. DOI: 10.2478/v10277-012-0019-4.

Hancock, Peter A. (2013). 'In search of vigilance: The problem of iatrogenically created psychological phenomena.' In: *American Psychologist* 68.2, pp. 97–109. ISSN: 1935-990X. DOI: 10.1037/a0030214.

Hancock, Peter A. et al. (2011). 'A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 53.5, pp. 517–527. ISSN: 0018-7208. DOI: 10.1177/0018720811417254.

Hart, Sandra G. (2006). 'NASA-task load index (NASA-TLX); 20 years later'. In: *Human Factors and Ergonomics Society Annual Meting* 50, pp. 904–908. ISSN: 1071-1813. DOI: 10.1037/e577632012-009.

Hart, Sandra G. and Lowell E. Staveland (1988). 'Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research'. In: *Advances in Psychology* 52, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9. URL: http://linkinghub.elsevier.com/retrieve/pii/S0166411508623869.

Hashim, Nor Hazlina and M. L. Jones (2007). 'Activity Theory: A framework for qualitative analysis'. In: *Qualitative Research*, pp. 3–5.

Hassenzahl, Marc (2005). 'The Thing and I: Understanding the Relationship Between User and Product'. In: *Funology*. Norwell, MA, USA: Kluwer Academic Publishers, pp. 31–42. ISBN: 1-4020-2966-7. DOI: 10.1007/1-4020-2967-5.

– (2006). 'Hedonic, Emotional, and Experiential Perspectives on Product Quality'. In: *Encyclopedia of Human Computer Interaction* 2000, pp. 266–272. ISSN: 10570810. DOI: 10.4018/978-1-59140-562-7.ch042.

Hassenzahl, Marc, Michael Burmester and Franz Koller (2003). 'AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität'. In: *Mensch & Computer 2003. Interaktion in Bewegung.* Ed. by J.Ziegler and G. Szwillus. Stuttgart, Leipzig: B.G. Teubner, pp. 187–196.

Hassenzahl, Marc, Sarah Diefenbach and Anja Göritz (2010). 'Needs, affect, and interactive products - Facets of user experience'. In: *Interacting with Computers* 22.5, pp. 353–362. ISSN: 09535438. DOI: 10.1016/j.intcom.2010.04.002.

Havard Health Publishing (2011). *Understanding the stress response.* URL: https://www.health.harvard.edu/staying-healthy/understanding-the-stress-response (visited on 20/03/2019).

Head, Emma (2009). 'The ethics and implications of paying participants in qualitative research'. In: *International Journal of Social Research Methodology* 12.4, pp. 335–344. ISSN: 14645300. DOI: 10.1080/13645570802246724.

Healey, Christopher G. and James T. Enns (1999). 'Large datasets at a glance: combining textures and colors in scientific visualization'. In: *IEEE Transactions on Visualization and Computer Graphics* 5.2, pp. 145–167. ISSN: 10772626. DOI: 10.1109/2945.773807.

Healey, Christopher and James Enns (2012). 'Attention and visual memory in visualization and computer graphics'. In: *IEEE Transactions on Visualization and Computer Graphics* 18.7, pp. 1170–1188. ISSN: 10772626. DOI: 10.1109/TVCG.2011.127.

Heimgärtner, Rüdiger, Alkesh Solanki and Helmut Windl (2017). 'Cultural User Experience in the Car – Toward a Standardized Systematic Intercultural Agile UI/UX Design Process'. In: *Automotive User Interfaces: Creating Interactive Experiences in the Car*, pp. 143–184. ISBN: 978-3-319-49447-0. DOI: 10.1007/978-3-319-49448-7.

Helldin, Tove et al. (2013). 'Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving'. In: *Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '13* 5, pp. 210–217. DOI: 10.1145/2516540.2516554.

Hergeth, Sebastian et al. (2016). 'Keep Your Scanners Peeled: Gaze Behavior as a Measure of Automation Trust During Highly Automated Driving'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 58.3, pp. 509–519. ISSN: 0018-7208. DOI: 10.1177/0018720815625744.

Hern, Alex (2016). *Germany calls on Tesla to drop 'Autopilot' branding.* URL: https://www.theguardian.com/technology/2016/oct/17/germany-calls-on-tesla-to-drop-autopilot-branding (visited on 19/03/2017).

Hills, Brian L. (1980). 'Vision , visibility , and perception in driving'. In: *Perception* 9, pp. 183–216.

Ho, Cristy, Nick Reed and Charles Spence (2006). 'Multisensory in-Car Warning Signais for Collision Avoidance'. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56.1, pp. 1107–1114. ISSN: 1071-1813. DOI: 10.1518/001872007X249965.

Ho, Cristy and Charles Spence (2005). 'Assessing the effectiveness of various auditory cues in capturing a driver's visual attention'. In: *Journal of Experimental Psychology: Applied* 11.3, pp. 157–174. ISSN: 1076-898X. DOI: 10.1037/1076-898X.11.3.157.

Ho, Cristy, Hong Z. Tan and Charles Spence (2005). 'Using spatial vibrotactile cues to direct visual attention in driving scenes'. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 8.6, pp. 397–412. ISSN: 13698478. DOI: 10.1016/j.trf.2005.05.002.

Hoeger, Reiner et al. (2011). 'The future of driving - HAVEit Final Report (D61.1)'. In: p. 358. URL: www.haveit-eu.org.

Hoff, Kevin Anthony and Masooda Bashir (2015). 'Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57.3, pp. 407–434. ISSN: 0018-7208. DOI: 10.1177/0018720814547570.

Hoffmann, Martin, Michael Wittke and Jorg Hahner (2009). 'Design and Evaluation of a Notification System for Alarm Management in Distributed Vision Networks'. In: *2009 29th IEEE International Conference on Distributed Computing Systems Workshops*, pp. 426–431. ISSN: 1545-0678. DOI: 10.1109/ICDCSW.2009.11.

Hollands, Justin G. and Ian Spence (1998). ‘Judging proportion with graphs: The summation model’. In: *Applied Cognitive Psychology* 12.June 1997, pp. 173–190. ISSN: 0888-4080. DOI: 10.1002/(SICI)1099-0720(199804)12:2<173::AID-ACP499>3.0.CO;2-K.

Holm, Sture (1979). ‘A Simple Sequentially Rejective Multiple Test Procedure’. In: *Scandinavian Journal of Statistics* 6.2, pp. 65–70.

Horberry, Tim et al. (2006). ‘Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance’. In: *Accident Analysis and Prevention* 38.1, pp. 185–191. ISSN: 00014575. DOI: 10.1016/j.aap.2005.09.007.

House of Lords Select Committee on Artificial Intelligence (2018). *AI in the UK: ready, willing and able? Report of session 2017-2019 HL Paper 100.* URL: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf (visited on 26/10/2019).

Howell, David C. (2013). *Statistical methods for psychology.* 8th ed. [S.l.]: Wadsworth Publishing Company, p. 770. ISBN: 9781111840853.

Hsieh, Hsiu-Fang and Sarah E. Shannon (2005). ‘Three Approaches to Qualitative Content Analysis’. In: *Qualitative Health Research* 15.9, pp. 1277–1288. ISSN: 1891943X. DOI: 10.1177/1049732305276687.

Hubbard, John R. and Edward A. Workman (1998). *Handbook of Stress Medicine: An Organ System Approach.* Boca Raton, FL, US: CRC Press, p. 423. ISBN: 9780849325151.

Huber, Daniel E. and Christopher G. Healey (2005). ‘Visualizing data with motion’. In: *Proceedings of the IEEE Visualization Conference*, p. 67. DOI: 10.1109/VIS.2005.125.

Hughes, Charmayne M.L. et al. (2015). ‘The application of SHERPA (Systematic Human Error Reduction and Prediction Approach) in the development of compensatory cognitive rehabilitation strategies for stroke patients with left and right brain damage’. In: *Ergonomics* 58.1, pp. 75–95. ISSN: 13665847. DOI: 10.1080/00140139.2014.957735.

ISO (1998). *Ergonomic requirements for offics work with visual display terminals (VDTs) – Part 11: Guidance on usability (BS EN ISO 9241-11:1998).* URL: https://bsol.bsigroup.com/Bibliographic/BibliographicInfoData/000000000030005417.

– (2006). *Ergonomics of human-system interaction – Part 20: Accessibility guidelines for information/communication technology (ICT) equipment and services (ISO 9241-20:2008).*

– (2009). *Ergonomics of human-system interaction – Part 920: Guidance on tactile and haptic interactions (ISO 9241-920:2009).*

– (2010). *Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems (BS EN ISO 9241-210:2010).*

– (2011a). *Ergonomics of human-system interaction – Part 303: Requirements for electronic visual displays (ISO 9241-303:2011).*

ISO (2011b). *Ergonomics of human-system interaction – Part 910: Framework for tactile and haptic interaction (ISO 9241-910:2011).*

– (2017). *BSI Standards Publication Road vehicles — Motorcycles — Symbols for controls , indicators and tell-tales (BS ISO 2575:2010+A7:2017).*

– (2018). *Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts (ISO 9241-11:2018).*

Israel, Boris (2012). 'Potenziale eines kontaktanalogen Head-up Displays für den Serieneinsatz'. PhD thesis. Technical University Munich. URL: https://mediatum.ub.tum.de/doc/1172451/1172451.pdf.

Itti, Laurent et al. (2001). 'Computational modelling of visual attention.' In: *Nature reviews. Neuroscience* 2.3, pp. 194–203. ISSN: 1471-003X. DOI: 10.1038/35058500.

Jacob, Robert J. K. and Keith S. Karn (2003). 'Commentary on Section 4 – Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises'. In: *The Mind's Eye*. Ed. by Jukka Hyönä, Ralph Radach and Heiner Deubel. Amsterdam, The Netherlands: Elsevier, pp. 573–605. DOI: https://doi.org/10.1016/B978-044451020-4/50031-1.

Jamson, A. Hamish et al. (2013). 'Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions'. In: *Transportation Research Part C: Emerging Technologies* 30, pp. 116–125. ISSN: 0968090X. DOI: 10.1016/j.trc.2013.02.008.

Ji, Yong Gu, Kwangil Lee and Wonil Hwang (2011). 'Haptic Perceptions in the Vehicle Seat'. In: *Human Factors and Ergonomics in Manufacturing & Service Industries* 21.3, pp. 305–325. DOI: 10.1002/hfm.20235.

Jian, Jiun-Yin, Ann M. Bisantz and Colin G. Drury (2000). 'Foundations for an Empirically Determined Scale of Trust in Automated System'. In: *International Journal of Cognitive Ergonomics* 4.1, p. 53. ISSN: 10886362. DOI: 10.1207/S15327566IJCE0401_04.

Jones, Debra G. (2014). 'A Practical Perspective on the Utility of Situation Awareness'. In: *Journal of Cognitive Engineering and Decision Making* 9.1, pp. 98–100. ISSN: 1555-3434. DOI: 10.1177/1555343414554804.

Jordan, Patrick W. (1998). *An introduction to usability.* English. London: Taylor & Francis. ISBN: 0748407944.

Julesz, B. and J. R. Bergen (1983). 'Human Factors and Behavioral Science: Textons, The Fundamental Elements in Preattentive Vision and Perception of Textures'. In: *Bell System Technical Journal* 62.6, pp. 1619–1645. ISSN: 15387305. DOI: 10.1002/j.1538-7305.1983.tb03502.x.

Jung, Ralf and Tim Schwartz (2008). 'A new approach to design and evaluate ambient systems for instrumented environments'. In: *2008 IET 4th International Conference on Intelligent Environments*. Seattle, WA, USA: IET. DOI: 10.1049/cp:20081172.

Jung, Tirza, Christina Kaß et al. (2017). 'So what really is user experience? An experimental study of user needs and emotional responses as underlying constructs'. In: *Ergonomics* 60.12, pp. 1601–1620. ISSN: 13665847. DOI: 10.1080/00140139.2017.1341555.

Juola, James F. (2016). 'Theories of Focal and Peripheral Attention'. In: *Peripheral Interaction: Challenges and Opportunities for HCI in the Periphery of Attention*. Ed. by Saskia Bakker, Doris Hausen and Ted Selker. Basel, Switzerland: Springer, pp. 39–61. ISBN: 978-3-319-29521-3. DOI: 10.1007/978-3-319-29523-7.

Kahol, Kanav et al. (2006). 'Evaluating the Role of Visio-Haptic Feedback in Multimodal Interfaces through EEG Analysis'. In: *Augmented Cognition: Past, Present and Future*. January, pp. 289–296.

Kaid, L. L. (1989). 'Content Analysis'. In: *Measurement of Communication Behavior*. Ed. by P. Emmert and L. L. Barker. New York, NY, US: Longman, pp. 197–217. ISBN: 0582286743.

Kaptein, Nico, Jan Theeuwes and Richard Van Der Horst (2007). 'Driving Simulator Validity: Some Considerations'. In: *Transportation Research Record: Journal of the Transportation Research Board* 1550.November 2015, pp. 30–36. ISSN: 0361-1981. DOI: 10.3141/1550-05.

Katsuki, Fumi and Christos Constantinidis (2014). 'Bottom-up and top-down attention: Different processes and overlapping neural systems'. In: *Neuroscientist* 20.5, pp. 509–521. ISSN: 10894098. DOI: 10.1177/1073858413514136.

Kauer, M., M. Schreiber and R. Bruder (2010). 'How to conduct a car? A design example for maneuver based driver-vehicle interaction'. In: *IEEE Intelligent Vehicles Symposium, Proceedings*, pp. 1214–1221. ISSN: 1931-0587. DOI: 10.1109/IVS.2010.5548099.

Keller, David and Stephen Rice (2010). 'System-Wide versus Component-Specific Trust Using Multiple Aids'. In: *The Journal of General Psychology* 137.1, pp. 114–128. ISSN: 0022-1309. DOI: 10.1080/00221300903266713.

Kennedy, Marc C and Anthony O'Hagan (2001). 'Bayesian Calibration of Computer Models'. In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 63.3, pp. 425–464. ISSN: 13697412, 14679868.

Khan, Khalid S et al. (2003). 'Five steps to conducting a systematic review'. In: *Journal of the Royal Society of Medicine* 96. ISSN: 0141-0768. DOI: 10.1258/jrsm.96.3.118.

Kim, Changseb, Jongwon Park and Kunsoo Huh (2016). 'Target classification layer design via vehicle-to-vehicle communication'. In: *Proceedings of the Institution of Mechanical*

Engineers, Part D: Journal of Automobile Engineering 230.13, pp. 1849–1861. ISSN: 09544070. DOI: 10.1177/0954407016633551.

Kircher, Katja and Christer Ahlström (2009). 'Issues related to the driver distraction detection algorithm AttenD'. In: *1st International Conferences on Driver Distraction and Inattention.* Gothenburg, Sweden, pp. 1–15. ISBN: 9781409425854.

Kirschenbaum, Susan S. and James E. Arruda (1994). 'Effects of Graphic and Verbal Probability Information on Command Decision Making'. In: *Human Factors* 36.3, pp. 406–418. ISSN: 00187208. DOI: 10.1177/001872089403600302.

Kiureghian, Armen Der and Ove Ditlevsen (2009). 'Aleatoric or Epistemic? Does it matter?' In: *Structural Safety* 31.2, pp. 105–112. ISSN: 01674730. DOI: 10.1016/j.strusafe.2008.06.020.

Klauer, Sheila G., Thomas A. Dingus et al. (2006). *The Impact of Driver Inattention On Near Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data (DOT HS 810 594).* Tech. rep. April. NHTSA, p. 226.

Klauer, Sheila G., Feng Guo et al. (2014). 'Distracted driving and risk of road crashes among novice and experienced drivers'. In: *The New England Journal of Medicine* 350, pp. 54–59. DOI: 10.1056/NEJMsa1204142.

Koo, Jeamin et al. (2015). 'Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance'. In: *International Journal on Interactive Design and Manufacturing* 9.4, pp. 269–275. ISSN: 19552505. DOI: 10.1007/s12008-014-0227-2.

Körber, Moritz, Eva Baseler and Klaus Bengler (2018). 'Introduction matters: Manipulating trust in automation and reliance in automated driving'. In: *Applied Ergonomics* 66, pp. 18–31. ISSN: 18729126. DOI: 10.1016/j.apergo.2017.07.006.

Körber, Moritz and Klaus Bengler (2013). 'Measurement of momentary user experience in an automotive context'. In: *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '13*, pp. 194–201. ISSN: 1931-0587. DOI: 10.1145/2516540.2516555.

Körber, Moritz, Christian Gold et al. (2016). 'The influence of age on the take-over of vehicle control in highly automated driving'. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 39, pp. 19–32. ISSN: 13698478. DOI: 10.1016/j.trf.2016.03.002.

Kowler, Eileen (2009). 'Attention and Eye Movements'. In: *Encyclopedia of Neuroscience.* Ed. by Larry R. Squire. Elsevier, Academic Press, pp. 605–616. DOI: 10.1016/B978-008045046-9.01087-1.

Kumar, Minu, Janell D. Townsend and Douglas W. Vorhies (2015). 'Enhancing Consumers' Affection for a Brand Using Product Design'. In: *Journal of Product Innovation Management* 32.5, pp. 716–730. ISSN: 15405885. DOI: 10.1111/jpim.12245.

Kunze, Alexander (2018). *Heartbeat Animation*. URL: https://dribbble.com/shots/4136090-Heartbeat-Animation (visited on 21/05/2018).

Kunze, Alexander, Stephen J. Summerskill et al. (2017). 'Enhancing driving safety and user experience through unobtrusive and function-specific feedback'. In: *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. Oldenburg, Germany: ACM, pp. 183–189. ISBN: 9781450351515. DOI: 10.1145/3131726.3131762.

– (2018a). 'Augmented Reality Displays for Communicating Uncertainty Information in Automated Driving'. In: *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '18*, pp. 164–175. DOI: 10.1145/3239060.3239074.

– (2018b). 'Evaluation of Variables for the Communication of Uncertainties Using Peripheral Awareness Displays'. In: *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. Toronto, ON, Canada: ACM. DOI: 10.1145/3239092.3265958.

– (2018c). 'Preliminary Evaluation of Variables for Communicating Uncertainties Using a Haptic Seat'. In: *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. Toronto, ON, Canada: ACM. DOI: 10.1145/3239092.3265959.

– (2019a). 'Automation Transparency: Implications of Uncertainty Communication for Human-Automation Interaction and Interfaces'. In: *Ergonomics* 62.3, pp. 345–360. DOI: 10.1080/00140139.2018.1547842.

– (2019b). 'Conveying Uncertainties Using Peripheral Awareness Displays in the Context of Automated Driving'. In: *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '19*. Utrecht, Netherlands: ACM. DOI: https://doi.org/10.1145/3342197.3344537.

– (2019c). 'Function-Specific Uncertainty Communication in Automated Driving'. In: *Special Issue of the International Journal of Mobile Human Computer Interaction (IJMHCI): Recent Advances in Automotive User Interfaces and Interactive Vehicular Applications Research* 11.2, p. 23. DOI: 10.4018/IJMHCI.2019040105.

Kuutti, Kari (1996). 'Activity theory as a potential framework for human- computer interaction research'. In: *Context and Consciousness: Activity Theory and Human Computer Interaction*, pp. 17–44.

Kuznetsova, Alexandra, Per Bruun Brockhoff and Rune Haubo Bojesen Christensen (2018). *Tests in Linear Mixed Effects Models*. Version 3.0-1. URL: https://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf.

Kyriakidis, Miltos et al. (2017). 'A human factors perspective on automated driving'. In: *Theoretical Issues in Ergonomics Science* 0.0, pp. 1–27. ISSN: 1463-922X. DOI: 10.1080/1463922X.2017.1293187.

Lachenmayr, B. (2006). 'Gesichtsfeld und Verkehr'. In: *Ophthalmologe* 103, pp. 373–381. ISSN: 0941293X. DOI: 10.1007/s00347-005-1268-5.

Lang, Virginia A., Michael Keith and Andrew Kavie (1994). 'Design Characteristics for Foveal and Peripheral Tasks in Multi-Task Visual Displays'. In: *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*. SAGE. DOI: https://doi.org/10.1177/154193129403801524.

Large, David R. and Gary E. Burnett (2019). 'Life on the Road: Exposing Drivers' Tendency to Anthropomorphise In-Vehicle Technology'. In: *Advances in Intelligent Systems and Computing: 20th Congress of the International Ergonomics Association, IEA 2018*. Vol. 823. Florence, Italy, pp. 3–12. ISBN: 978-3-319-96097-5. DOI: 10.1007/978-3-319-96098-2.

Large, David R., Gary E. Burnett, Andrew Morris and Arun Muthumani (2017). 'A Longitudinal Simulator Study to Explore Drivers' Behaviour During Highly-Automated Driving'. In: *Road Safety and Simulation International Conference*. July. The Hague, NL. ISBN: 978-3-319-60440-4. DOI: 10.1007/978-3-319-60441-1.

Large, David R., Gary Burnett, Davide Salanitri et al. (2019). 'A longitudinal simulator study to explore drivers' behaviour in level 3 automated vehicles'. In: *Proceedings - 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019*, pp. 222–232. DOI: 10.1145/3342197.3344519.

Large, David R, Gary Burnett, Andrew Morris, Arun Muthumani and Rebecca Matthias (2017). 'Design Implications of Drivers' Engagement with Secondary Activities During Highly - Automated Driving – A Longitudinal Simulator Study'. In: *Road Safety & Simulation International Conference 2017* October, pp. 1–10.

Lauber, Felix, Claudius Böttcher and Andreas Butz (2014). 'You've Got the Look: Visualizing Infotainment Shortcuts in Head-Mounted Displays'. In: *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Seattle, WA, USA: ACM.

Law Commission of England and Wales and Scottish Law Commission (2018). *Automated Vehicles: A joint preliminary consultation paper*. URL: https://www.lawcom.gov.uk/project/automated-vehicles/ (visited on 26/10/2019).

LeCun, Yann, Yoshua Bengio and Geoffrey Hinton (2015). 'Deep Learning'. In: *Nature* 521.7553, pp. 436–444. ISSN: 0028-0836. DOI: 10.1038/nature14539.

Lee, John D. (2018). 'Perspectives on Automotive Automation and Autonomy'. In: *Journal of Cognitive Engineering and Decision Making* 12.1, pp. 53–57. ISSN: 21695032. DOI: 10.1177/1555343417726476.

Lee, John D., Daniel V. McGehee et al. (2002). 'Collision Warning Timing, Driver Distraction, and Driver Response to Imminent Rear-End Collisions in a High-Fidelity Driving Simulator'. In: 44.2, pp. 314–334.

Lee, John D. and Neville Moray (1992). 'Trust, Control Strategies and Allocation of Function in Human-Machine Systems'. In: *Ergonomics* 35.March, pp. 1243–1270. ISSN: 0014-0139. DOI: 10.1080/00140139208967392.

– (1994). 'Trust, self-confidence, and operators' adaptation to automation'. In: *International Journal of Human Computer Studies* 40, pp. 153–184.

Lee, John D. and Katrina A. See (2004). 'Trust in Automation: Designing for Appropriate Reliance'. In: *Human Factors* 46.1, pp. 50–80.

Leibowitz, Herschel W. et al. (1982). 'Implications of Recent Developments in Dynamic Spatial Orientation and Visual Resolution for Vehicle Guidance'. In: *Tutorials on Motion Perception*. Ed. by Alexander H Wertheim, Willem A Wagenaar and Herschel W Leibowitz. Boston, MA: Springer US, pp. 231–260. ISBN: 978-1-4613-3569-6. DOI: 10.1007/978-1-4613-3569-6_8.

Leontyev, Aleksei N. (1981). *The Development of Mind*. Moscow: Progress Press.

Liljamo, Timo, Heikki Liimatainen and Markus Pöllänen (2018). 'Attitudes and concerns on automated vehicles'. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 59, pp. 24–44. ISSN: 13698478. DOI: 10.1016/j.trf.2018.08.010.

Lin, Rui, Liang Ma and Wei Zhang (2018). 'An interview study exploring Tesla drivers' behavioural adaptation'. In: *Applied Ergonomics* 72.April, pp. 37–47. ISSN: 00036870. DOI: 10.1016/j.apergo.2018.04.006.

Löcken, Andreas et al. (2013). 'AmbiCar: Towards an in-vehicle ambient light display'. In: *AutomotiveUI 2013* October, pp. 107–108.

Loecken, Andreas, Wilko Heuten and Susanne Boll (2015). 'Supporting Lane Change Decisions with Ambient Light'. In: *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Nottingham, UK: ACM, pp. 204–211. ISBN: 9781450337366. DOI: 10.1145/2799250.2799259.

Löper, Christian et al. (2006). 'Path planning, vehicle guidance and haptic interaction for a driver assistance system for lateral guidance'. In: *VDI-Berichte* Nr. 1960, pp. 189–205.

Louw, Tyron, Jonny Kuo et al. (2019). 'Engaging in NDRTs affects drivers ' responses and glance patterns after silent automation failures'. In: *Transportation Research Part F:*

Psychology and Behaviour 62.March, pp. 870–882. ISSN: 1369-8478. DOI: 10.1016/j.trf.2019.03.020.

Louw, Tyron, Gustav Markkula et al. (2017). 'Coming back into the loop: Drivers' perceptual-motor performance in critical events after automated driving'. In: *Accident Analysis and Prevention* 108.September, pp. 9–18. ISSN: 00014575. DOI: 10.1016/j.aap.2017.08.011.

Louw, Tyron and Natasha Merat (2017). 'Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation'. In: *Transportation Research Part C* 90.November, pp. 1–18. ISSN: 0968090X. DOI: 10.1016/j.trc.2017.01.001.

Lu, Zhenji et al. (2016). 'Human factors of transitions in automated driving: A general framework and literature survey'. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 43, pp. 183–198. ISSN: 13698478. DOI: 10.1016/j.trf.2016.10.007.

MacEachren, Alan M. (1992). 'Visualizing uncertain information'. In: *Cartographic Perspective* 13, pp. 10–19. ISSN: 1048-9053. DOI: 10.1.1.62.285.

– (2004). *How Maps Work: Representation, Visualization, and Design*. Guilford Press, p. 513. ISBN: 978-1572300408.

MacEachren, Alan M., Anthony Robinson et al. (2005). 'Visualizing geospatial information uncertainty: What we know and what we need to know'. In: *Cartography and Geographic Information ScienceInformation Science* 32.3, pp. 139–160. ISSN: 1523-0406. DOI: 10.1559/1523040054738936.

MacEachren, Alan M., Robert E. Roth et al. (2012). 'Visual Semiotics and Uncertainty Visualisation: An Empirical Study'. In: *IEEE Transactions on Visualization and Computer Graphics*. Vol. 18. 12. IEEE, pp. 2496–2505. DOI: 10.1109/TVCG.2012.279.

MacKison, D., W. L. Wrieden and A. S. Anderson (2010). 'Validity and reliability testing of a short questionnaire developed to assess consumers use, understanding and perception of food labels'. In: *European Journal of Clinical Nutrition* 64.2, pp. 210–217. ISSN: 09543007. DOI: 10.1038/ejcn.2009.126.

Madigan, Ruth, Tyron Louw and Natasha Merat Institute (2018). 'The effect of varying levels of vehicle automation on drivers' lane changing behaviour'. In: *PLoS ONE* 13.2, pp. 1–17. DOI: https://doi.org/10.1371/journal.pone.0192190.

Manca, Laura, Joost C.F. de Winter and Riender Happee (2015). 'Visual Displays for Automated Driving: a Survey'. In: *Workshop on Adaptive Ambient In-Vehicle Displays and Interactions - AutomotiveUI '15*, pp. 1–5. ISBN: 9781450337366.

Mankoff, Jennifer and Anind K. Dey (2003). 'From Conception to Design'. In: *Public and Situated Displays. The Kluwer International series on Computer Supported Cooperative Work*. Ed. by Kenton O'Hara et al. Springer. ISBN: 978-94-017-2813-3. DOI: 10.1007/978-94-017-2813-3_9.

Mankoff, Jennifer, Anind K. Dey et al. (2003). 'Heuristic Evaluation of Ambient Displays'. In: *Proceedings of the SIGCHI conference on Human factors in computing systems* 5, pp. 169–176. ISSN: 1941-6237. DOI: 10.1145/642611.642642.

Manzey, Dietrich, Juliane Reichenbach and Linda Onnasch (2012). 'Human Performance Consequences of Automated Decision Aids: The Impact of Degree of Automation and System Experience'. In: *Journal of Cognitive Engineering and Decision Making* 6.1, pp. 57–87. ISSN: 1555-3434. DOI: 10.1177/1555343411433844.

Matthews, Tara, Anind K. Dey et al. (2004). 'A Toolkit for Managing User Attention in Peripheral Displays'. In: *UIST '04*. Santa Fe, New Mexico, USA: ACM, pp. 247–256. ISBN: 1581139578.

Matthews, Tara, Gary Hsieh and Jennifer Mankoff (2009). 'Evaluating Peripheral Displays'. In: *Awareness Systems. Human-Computer Interaction Series.* Ed. by Panos Markopoulos, Boris De Ruyter and Wendy Mackay. London: Springer, pp. 447–472. ISBN: 978-1-84882-476-8. DOI: 10.1007/978-1-84882-477-5_19.

Matthews, Tara, Tye Rattenbury and Scott Carter (2007). 'Defining, Designing, and Evaluating Peripheral Displays - An Analysis Using Activity Theory'. In: *Human–Computer Interaction* 22, pp. 221–261. ISSN: 0737-0024. DOI: 10.1080/07370020701307997.

Matviienko, Andrii et al. (2015). 'Deriving design guidelines for ambient light systems'. In: *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia - MUM '15* 30-Novembe.Mum, pp. 267–277. DOI: 10.1145/2836041.2836069.

May, P., Robert Molloy and Raja Parasuraman (1993). 'Effects of automation reliability and failure rate on monitoring performance in a multi-task environment'. In: *Paper presented at the annual meeting of the Human Factors Society, Santa Monica, CA* 37.

McCandless, David, Pearl Doughty-White and Miriam Quick (2015). *Codebases: Million of lines of code.* URL: http://www.informationisbeautiful.net/visualizations/million-lines-of-code/ (visited on 25/02/2017).

McCraty, Rollin and Fred Shaffer (2015). 'Heart rate variability: New perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk'. In: *Global Advances In Health and Medicine* 4.1, pp. 46–61. ISSN: 21649561. DOI: 10.7453/gahmj.2014.073.

McCrickard, D. Scott and C. M. Chewar (2003). 'Attuning Notification Design to User Goals and Attention Costs'. In: *Communications of the ACM* 46.3, pp. 67–72.

McCrickard, D. Scott, C. M. Chewar et al. (2003). 'A model for notification systems evaluation – assessing user goals for multitasking activity'. In: *ACM Transactions on Computer-Human Interaction* 10.4, pp. 312–338. ISSN: 10730516. DOI: 10.1145/966930.966933.

McGuirl, John M. and Nadine B. Sarter (2003). 'How are we doing?: Presenting System Confidence Information to Support Trust Calibration and Adaptive Function Allocation'. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 47.3, pp. 538–542. ISSN: 1071-1813. DOI: 10.1177/154193120304700360.

– (2006). 'Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information'. In: *Human Factors* 48.4, pp. 656–665. ISSN: 00187208. DOI: 10.1518/001872006779166334.

McKendrick, Ryan D. and Erin Cherry (2018). 'A Deeper Look at the NASA TLX and Where It Falls Short'. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 62.1, pp. 44–48. ISSN: 1541-9312. DOI: 10.1177/1541931218621010.

McKenna, Frank P., Robert A. Stanier and Clive Lewis (1991). 'Factors underlying illusory self- assessment of driving skill in males and females'. In: *Accident Analysis and Prevention* 23.1, pp. 45–52. ISSN: 0001-4575. DOI: 10.1016/0001-4575(91)90034-3.

McMains, Stephanie A. and Sabine Kastner (2009). 'Visual Attention'. In: *Encyclopedia of Neuroscience*. Ed. by Marc D. Binder, Nobutaka Hirokawa and Uwe Windhorst. Berlin, Heidelberg: Springer. ISBN: 978-3-540-29678-2. DOI: 10.1007/978-3-540-29678-2_6344.

Mengelkoch, Robert F., Jack A. Adams and Charles A. Gainer (1971). 'The forgetting of instrument flying skills'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 13.5, pp. 397–405. DOI: 10.1177/001872087101300502.

Mercado, Joseph E. et al. (2016). 'Intelligent Agent Transparency in Human–Agent Teaming for Multi-UxV Management'. In: *Human Factors* 58.3, pp. 401–415. ISSN: 0018-7208. DOI: 10.1177/0018720815621206.

Met Office UK (2018). *What is fog?* URL: https://www.metoffice.gov.uk/learning/clouds/fog (visited on 25/05/2018).

Metzger, Ulla and Raja Parasuraman (2005). 'Automation in Future Air Traffic Management: Effects of Decision Aid Reliability on Controller Performance and Mental Workload'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 47.1, pp. 1–15.

Meyer, Gereon (2015). *Road Vehicle Automation 2*. ISBN: 9783319059891.

Meyer, Joachim (2001). 'Effects of Warning Validity and Proximity on Responses to Warnings'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 43.4, pp. 563–572. ISSN: 0018-7208. DOI: 10.1518/001872001775870395.

– (2004). 'Conceptual Issues in the Study of Dynamic Hazard Warnings'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46.2, pp. 196–204. ISSN: 0018-7208. DOI: 10.1518/hfes.46.2.196.37335.

Michon, John A. (1985). ‘A critical view of driver behavior models: what do we know, what should we do?’ In: *Human Behavior and Traffic Safety*. Ed. by Leonard Evans and Richard C. Swing. New York: Plenum Press, pp. 485–520. ISBN: 978-1-4612-9280-7. DOI: 10.1007/978-1-4613-2173-6.

Miglani, Abhijai, Cyriel Diels and Jacques Terken (2016). ‘Compatibility between Trust and Non-Driving Related Tasks in UI Design for Highly and Fully Automated Driving’. In: *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct - Automotive'UI 16*, pp. 75–80. DOI: 10.1145/3004323.3004331.

Miller, Chris (2005). *Using Delegation as an Architecture for Adaptive Automation*. Tech. rep. December. Air Force Research Laboratory.

Miller, David, Mishel Johns et al. (2016). ‘Behavioral Measurement of Trust in Automation: The Trust Fall’. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 60.1, pp. 1849–1853. ISSN: 1541-9312. DOI: 10.1177/1541931213601422.

Miller, David, Annabel Sun et al. (2015). ‘Distraction Becomes Engagement in Automated Driving’. In: *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting2*, pp. 2–6. ISSN: 1541-9312. DOI: 10.1177/1541931215591362.

Minderhoud, Michiel M. and Piet H.L. Bovy (2001). ‘Extended time-to-collision measures for road traffic safety assessment’. In: *Accident Analysis and Prevention* 33.1, pp. 89–97. ISSN: 00014575. DOI: 10.1016/S0001-4575(00)00019-1.

Minge, Michael (2018a). *Evaluation of Interactive Products*. URL: http://mecue.de/Homepage%20Content/english/meCUE_PV.pdf (visited on 12/06/2018).

– (2018b). *Guide to using the meCUE 2.0 questionnaire*. URL: http://mecue.de/Homepage%20Content/english/meCUE_EV.pdf (visited on 12/06/2018).

Minge, Michael and Manfred Thuering (2016). ‘The meCUE Questionaire: A Modular Tool for Measuring User Experience’. In: *Advances in Ergonomics Modeling, Usability & Special Populations. Proceedings of the 7th Applied Human Factors and Ergonomics Society Conference 2016* 486.January. ISSN: 02683768. DOI: 10.1007/978-3-319-41685-4.

Minge, Michael, Manfred Thüring and Ingmar Wagner (2016). ‘Developing and Validating an English Version of the meCUE Questionnaire for Measuring User Experience’. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 60.1, pp. 2063–2067. ISSN: 1541-9312. DOI: 10.1177/1541931213601468.

Mirnig, Alexander G. et al. (2019). ‘The Insurer's Paradox : About Liability , the Need for Accident Data, and Legal Hurdles for Automated Driving’. In: *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular*

Applications - AutomotiveUI '19. Utrecht, Netherlands: ACM, pp. 113–122. DOI: 10.1145/3342197.3344540.

Miyamae, Masakazu et al. (2008). 'Design guidelines of a context-aware notification system for nursing'. In: *Proceedings of the 2nd International Conference on Pervasive Computing Technologies for Healthcare 2008, PervasiveHealth*, pp. 156–159. DOI: 10.1109/PCTHEALTH.2008.4571057.

Mok, Brian et al. (2017). 'Tunneled In: Drivers with Active Secondary Tasks Need More Time to Transition from Automation'. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, pp. 2840–2844. DOI: 10.1145/3025453.3025713.

Molich, Rolf and Jakob Nielsen (1990). 'Improving a human-computer dialogue'. In: *Communications of the ACM* 33.3, pp. 338–348. ISSN: 00010782. DOI: 10.1145/77481.77486.

Molloy, Robert and Raja Parasuraman (1996). 'Monitoring an Automated System for a Single Failure: Vigilance and Task Complexity Effects'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 38.2, pp. 311–322. ISSN: 0018-7208. DOI: 10.1177/001872089606380211.

Moray, Neville and Toshiyuki Inagaki (2000). 'Attention and complacency'. In: *Theoretical Issues in Ergonomics Science* 1.4, pp. 354–365. ISSN: 1463-922X. DOI: 10.1080/14639220052399159.

Morf, Martin E and Wolfgang G Weber (2000). 'I/O Psychology and the bridging of A. N. Leont'ev's activity theory.' In: *Canadian Psychology/Psychologie canadienne* 41.2, pp. 81–93. ISSN: 1878-7304(Electronic),0708-5591(Print). DOI: 10.1037/h0088234.

Morgan, David L. (1993). 'Qualitative Content Analysis: A Guide to Paths Not Taken'. In: *Qualitative Health Research1* 3.1, pp. 112–121.

Morgan, David L. and Heather Guevara (2008). 'The SAGE Encyclopedia of Qualitative Methods'. In: *The SAGE Encyclopedia of Qualitative Research Methods*. Ed. by Lisa M. Given. Thousand Oaks: SAGE Publications, Inc., p. 470. DOI: http://dx.doi.org/10.4135/9781412963909.n238.

Morris, Drew M., Jason M. Erno and June J. Pilcher (2017). 'Electrodermal response and automation trust during simulated self-driving car use'. In: *Proceedings of the Human Factors and Ergonomics Society*, pp. 1759–1762. ISBN: 9780945289531. DOI: 10.1177/1541931213601921.

Morrison, Joel L. (1974). 'A theoretical framework for cartographic generalization with the emphasis on the process of symbolization'. In: *International Yearbook of Cartography* 14, pp. 115–127.

Most, Steven B. and Daniel J. Simons (1999). 'Sustained Inattentional Blindness: Dynamic Events'. In: *Perception* 28, pp. 1059–1074.

Mueller, Heiko et al. (2014). 'Sparkle : An Ambient Light Display for Dynamic Off-Screen Points of Interest'. In: *NordiCHI '14 Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*. Helsinki, Finland: ACM, pp. 51–60. ISBN: 9781450325424.

Muir, Bonnie M. (1987). 'Trust between humans and machines, and the design of decision aids'. In: *International Journal of Machine Studies* 27, pp. 527–539.

Muir, Bonnie M. and Neville Moray (1996). 'Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation'. In: *Ergonomics* 39.3, pp. 429–60. ISSN: 0014-0139. DOI: 10.1080/00140139608964474.

Nass, Clifford and Youngme Moon (2000). 'Machines and Mindlessness: Social Responses to Computers'. In: *Journal of Social Issues* 56.1, pp. 81–103. ISSN: 0022-4537. DOI: 10.1111/0022-4537.00153.

National Center for Statistics and Analysis (2019). 'Driver electronic device use in 2017'. In: *Traffic Safety Facts Research Note. Report No. DOT HS 812 665* 2017.December, pp. 1–8. DOI: 10.1109/FTCS.1998.689468. URL: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812665.

National Highway Traffic Safety Administration (2013). 'National Highway Traffic Safety Administration Preliminary Statement of Policy Concerning Automated Vehicles'. In: *National Highway Traffic Safety Administration*, p. 14.

National Research Council (2014). *Complex Operational Decision Making in Networked Systems of Humans and Machines: A Multidisciplinary Approach*. ISBN: 9780309307703. DOI: 10.17226/18844.

National Research Council (US) Committee on Vision (1985). *Emergent techniques for assessment of visual performance*. Washington, DC: National Academies Press (US). ISBN: 030959166X. URL: https://www.ncbi.nlm.nih.gov/books/NBK219039/.

Naujoks, Frederik, Yannick Forster et al. (2016). 'Speech improves human-automation cooperation in automated driving'. In: *Mensch und Computer 2016 – Workshopband* September. DOI: 10.18420/muc2016-ws08-0007.

– (2017). 'A Human-Machine Interface for Cooperative Highly Automated Driving'. In: *Advances in Human Aspects of Transportation*, pp. 585–595. DOI: 10.1007/978-3-319-41682-3.

Naujoks, Frederik, Christoph Mai and Alexandra Neukum (2014). 'The Effect of Urgency of Take-Over Requests During Highly Automated Driving Under Distraction Conditions'. In: *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE*. Ed. by T. Ahram, W. Karwowski and T. Marek. July. Krakow, pp. 2099–2106.

Naujoks, Frederik, Christian Purucker and Alexandra Neukum (2016). 'Secondary task engagement and vehicle automation - Comparing the effects of different automation levels in an on-road experiment'. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 38, pp. 67–82. ISSN: 13698478. DOI: 10.1016/j.trf.2016.01.011.

Neale, Vicky L. and Thomas A. Dingus (1998). 'Human factors issues for automated highways systems (AHS)'. In: *Intelligent Transportation Systems Journal* 4.1-2, pp. 110–112. DOI: 10.1080/10248079808903740.

Newman, William M. (1997). 'Better or Just Different? On the Benefits of Designing Interactive Systems in terms of Critical Parameters'. In: *DIS '97 Proceedings of the 2nd conference on Designing interactive systems: processes, practices, methods, and techniques* August, pp. 239–245. DOI: 10.1145/263552.263615.

Neyedli, Heather F., Justin G. Hollands and Greg A. Jamieson (2011). 'Beyond Identity: Incorporating System Reliability Information Into an Automated Combat Identification System'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 53.4, pp. 338–355. ISSN: 0018-7208. DOI: 10.1177/0018720811413767.

Nguyen, Anh, Jason Yosinski and Jeff Clune (2015). 'Deep Neural Networks are Easily Fooled'. In: *Computer Vision and Pattern Recognition, 2015 IEEE Conference on*, pp. 427–436. ISSN: 1875-7855. DOI: 10.1109/CVPR.2015.7298640.

Nguyen, Thanh, Chee Peng Lim et al. (2019). 'A Review of Situation Awareness Assessment Approaches in Aviation Environments'. In: *IEEE Systems Journal*. DOI: 10.1109/JSYST.2019.2918283.

NHTSA (1995). *Human factors aspects of using head up displays in automobiles : A review of the literature (DOT HS 808 320)*. Tech. rep. August, p. 84.

– (2008). *National Motor Vehicle Crash Causation Survey Report to Congress (DOT HS 811 059)*. Tech. rep. July, p. 47.

– (2018). *Teen Distracted Driver Data (DOT HS 812 667)*. Tech. rep. URL: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812504.

Nielsen, Jakob (1994a). 'Enhancing the explanatory power of usability heuristics'. In: *Conference companion on Human factors in computing systems - CHI '94*, p. 210. ISSN: 0897916506. DOI: 10.1145/259963.260333.

– (1994b). *Usability Inspection Methods*. New York, NY: John Wiley & Sons. ISBN: 0-471-01877-5.

Nielsen, Jakob and Thomas K. Landauer (1993). 'A Mathematical Model of the Finding of Usability Problems'. In: *CHI '93 Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*. Amsterdam, The Netherlands: ACM, pp. 206–213. DOI: 10.1145/169059.169166.

Nielsen, Jakob and Rolf Molich (1990). 'Heuristic Evaluation of user interfaces'. In: *CHI '90 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* April, pp. 249–256. ISSN: 1942-597X. DOI: 10.1145/97243.97281.

Noah, Brittany E., Thomas M. Gable and Bruce N. Walker (2016). 'Ordinal Magnitude Scaling for Automated Lane Keeping Displays Abstract'. In: *8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '16), October 24–26, 2016*.

Noah, Brittany E., Thomas. M. Gable, Shao-Yu Chen et al. (2017). 'Development and Preliminary Evaluation of Reliability Displays for Automated Lane Keeping'. In: *Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '17*. Oldenburg, Germany: ACM, pp. 202–208. ISBN: 9781450351508. DOI: 10.1145/3122986.3123007.

Noguchi, Kimihiro et al. (2012). 'nparLD: An R Software Package for the Nonparametric Analysis of Longitudinal Data in Factorial Experiments'. In: *Journal of Statistical Software* 50.12, pp. 1–23. URL: http://www.jstatsoft.org/v50/i12/.

Norman, Donald A. (1989). 'The "Problem" of Automation: Inappropriate Feedback and Interaction, Not "Overautomation"'. In: *Philosphiccal Transactions of the Royal Society of London* B.

– (2013). *The Design of Everyday Things*. Revised an. Basic Books, p. 272. ISBN: 0465067107. DOI: 10.1002/hfm.20127.

– (2015). 'The Human Side of Automation'. In: *Road Vehicle Automation 2*. Springer, pp. 73–79. DOI: DOI10.1007/978-3-319-19078-5_7.

O'Neal, Wesley T. et al. (2016). 'Reference Ranges for Short-Term Heart Rate Variability Measures in Individuals Free of Cardiovascular Disease: The Multi-Ethnic Study of Atherosclerosis (MESA)'. In: *Journal of Electrocardiology* 49.5, pp. 686–690. DOI: 10.1016/j.jelectrocard.2016.06.008.

OECD (2015). *Automated and Autonomous Driving: Regulation under uncertainty*. URL: https://www.itf-oecd.org/sites/default/files/docs/15cpb_autonomousdriving.pdf (visited on 26/10/2019).

Ongenae, F. et al. (2012). 'User-driven design of a context-aware application: an ambient-intelligent nurse call system'. In: *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. San Diego, California, USA: IEEE, pp. 205–210. ISBN: 9781936968435. DOI: 10.4108/icst.pervasivehealth.2012.248699.

Ou, Li Chen et al. (2012). 'Age effects on colour emotion, preference, and harmony'. In: *Color Research and Application* 37.2, pp. 92–105. ISSN: 03612317. DOI: 10.1002/col.20672.

Oxford Dictionaries (2016a). *Automation*. URL: https://en.oxforddictionaries.com/definition/automation (visited on 04/11/2016).

– (2016b). *Vigilance*. URL: https://en.oxforddictionaries.com/definition/vigilance (visited on 04/11/2016).

– (2018a). *Display*. URL: https://en.oxforddictionaries.com/definition/display (visited on 14/06/2018).

– (2018b). *Heuristic*. URL: https://en.oxforddictionaries.com/definition/heuristic (visited on 04/06/2018).

– (2018c). *Notify*. URL: https://en.oxforddictionaries.com/definition/notify (visited on 14/06/2018).

– (2018d). *System*. URL: https://en.oxforddictionaries.com/definition/system (visited on 14/06/2018).

Pak, Richard et al. (2012). 'Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults'. In: *Ergonomics* 55.9, pp. 1059–1072. ISSN: 00140139. DOI: 10.1080/00140139.2012.691554.

Pang, Alex T., Craig M. Wittenbrink and Suresh K. Lodha (1997). 'Approaches to uncertainty visualization'. In: *The Visual Computer* 13.8, pp. 370–390. ISSN: 01782789. DOI: 10.1007/s003710050111.

Parasuraman, Raja (1986). 'Vigilance, monitoring, and search'. In: *Handbook of perception and human performance. Vol. 2: Cognitive processes and performance*. Ed. by K. Boff, L. Kaufmann and J. Thomas. New York, NY, US: Wiley, pp. 43.1–43.39.

Parasuraman, Raja and Dietrich Manzey (2010). 'Complacency and bias in human use of automation: an attentional integration.' In: *Human factors* 52, pp. 381–410. ISSN: 0018-7208. DOI: 10.1177/0018720810376055.

Parasuraman, Raja, Robert Molloy and Indramani L. Singh (1993). 'Performance Consequences of Automation-Induced 'Complacency''. In: *The International Journal of Aviation Psychology* 3.1, pp. 1–23. DOI: 10.1207/s15327108ijap0301_1.

Parasuraman, Raja and Victor Riley (1997). 'Humans and Automation: Use, Misuse, Disuse, Abuse'. In: *Human Factors The Journal of the Human Factors and Ergonomics Society* 39.2, pp. 230–253.

Parasuraman, Raja, Thomas B. Sheridan and Christopher D. Wickens (2000). 'A model for types and levels of human interaction with automation.' In: *IEEE transactions on systems, man, and cybernetics. Part A, Systems and humans : a publication of the IEEE Systems, Man, and Cybernetics Society* 30.3, pp. 286–297. ISSN: 1083-4427. DOI: 10.1109/3468.844354.

– (2008). 'Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs'. In: *Journal of Cognitive En-*

gineering and Decision Making 2.2, pp. 140–160. ISSN: 1555-3434. DOI: 10.1518/155534308X284417.

Parrish, Kristen (2015). 'Under the Hood: Tech Behind Autonomous Vehicles'. In: *SXSW*. Austin, Texas: Texas Instruments.

Patterson, Robert Earl (2017). 'Intuitive Cognition and Models of Human-Automation Interaction'. In: *Human Factors* 59.1, pp. 101–115. ISSN: 15478181. DOI: 10.1177/0018720816659796.

Patterson, Robert Earl and Robert G. Eggleston (2017). 'Intuitive Cognition'. In: *Journal of Cognitive Engineering and Decision Making* 11.1, pp. 5–22. ISSN: 21695032. DOI: 10.1177/1555343416686476.

Payre, William, Julien Cestac and Patricia Delhomme (2016). 'Fully Automated Driving: Impact of Trust and Practice on Manual Control Recovery'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society*, pp. 229–241. ISSN: 0018-7208. DOI: 10.1177/0018720815612319. URL: http://hfs.sagepub.com/content/early/2015/11/25/0018720815612319.abstract.

Pettersson, Ingrid et al. (2018). 'A Bermuda Triangle? - A Review of Method Application and Triangulation in User Experience Evaluation'. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pp. 1–16. DOI: 10.1145/3173574.3174035.

Pfannmüller, Lisa et al. (2015). 'A Comparison of Display Concepts for a Navigation System in an Automotive Contact Analog Head-up Display'. In: *Procedia Manufacturing* 3, pp. 2722–2729. ISSN: 23519789. DOI: 10.1016/j.promfg.2015.07.678.

Pfleging, Bastian and Albrecht Schmidt (2015). '(Non-) Driving-Related Activities in the Car: Defining Driver Activities for Manual and Automated Driving'. In: *Chi '15*, pp. 5–8.

Piechulla, Walter et al. (2003). 'Reducing drivers' mental workload by means of an adaptive man-machine interface'. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 6.4, pp. 233–248. ISSN: 13698478. DOI: 10.1016/j.trf.2003.08.001.

Politis, Ioannis (2016). 'Effects of Modality, Urgency and Situation on Responses to Multimodal Warnings for Drivers'. PhD thesis. University of Glasgow.

Politis, Ioannis, Stephen Brewster and Frank Pollick (2013). 'Evaluating multimodal driver displays under varying situational urgency'. In: *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '13*. Eindhoven, Netherlands: ACM, pp. 92–99. ISBN: 9781450324731. DOI: 10.1145/2516540.2516543.

– (2014). 'Evaluating multimodal driver displays under varying situational urgency'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Toronto, ON, Canada: ACM, pp. 4067–4076. ISBN: 9781450324731. DOI: 10.1145/2556288.2556988.

Politis, Ioannis, Stephen Brewster and Frank Pollick (2015a). 'Language-based multimodal displays for the handover of control in autonomous cars'. In: *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '15*. Nottingham UK: ACM, pp. 3–10. ISBN: 9781450337366. DOI: 10.1145/2799250.2799262.

– (2015b). 'To Beep or Not to Beep?: Comparing Abstract versus Language-Based Multimodal Driver Displays'. In: *Proceedings of the ACM CHI'15 Conference on Human Factors in Computing Systems*. Seoul, Republic of Korea: ACM, pp. 3971–3980. ISBN: 9781450331456. DOI: 10.1145/2702123.2702167.

Posner, Michael I. (1980). 'Orienting of attention'. In: *Quarterly Journal of Experimental Psychology* 21.1, pp. 3–25. ISSN: 0033-555X. DOI: 10.1080/00335558008248231. arXiv: arXiv:1011.1669v3.

Pousman, Zachary and John Stasko (2006). 'A Taxonomy of Ambient Information Systems: Four Patterns of Design'. In: *Proceedings of the working conference on Advanced visual interfaces - AVI '06*, p. 67. DOI: 10.1145/1133265.1133277.

Precision Microdrives Limited (2017). *Product Data Sheet 9mm Vibration Motor – 25mm Type*. Model: 307-103.005.

– (2018). *How To Drive A Vibration Motor With Arduino And Genuino*. URL: https://www.precisionmicrodrives.com/content/how-to-drive-a-vibration-motor-with-arduino-and-genuino/ (visited on 12/10/2018).

– (2019). *AB-029 : Vibration Motors – Voltage Vs Frequency Vs Amplitude*. URL: https://www.precisionmicrodrives.com/content/ab-029-vibration-motors-voltage-vs-frequency-vs-amplitude/ (visited on 02/02/2019).

Proctor, Robert W and Kim-Phuong L Vu (2012). 'Human Information Processing'. In: *Encyclopedia of the Sciences of Learning*. Ed. by Norbert M Seel. Boston, MA: Springer US, pp. 1458–1460. ISBN: 978-1-4419-1428-6. DOI: 10.1007/978-1-4419-1428-6_722.

R Core Team (2019). *R: A Language and Environment for Statistical Computing [Version 3.5.3]*. Vienna, Austria. URL: http://www.r-project.org/.

Rasshofer, R. H., M. Spies and H. Spies (2011). 'Influences of weather phenomena on automotive laser radar systems'. In: *Advances in Radio Science* 9, pp. 49–60. ISSN: 16849965. DOI: 10.5194/ars-9-49-2011.

Rayner, Keith (2009). 'Eye movements and attention in reading, scene perception, and visual search'. In: *Quarterly Journal of Experimental Psychology* 62.8, pp. 1457–1506. DOI: 10.1080/17470210902816461.

Rice, Stephen and Kasha Geels (2010). 'Using system-wide trust theory to make predictions about dependence on four diagnostic aids'. In: *Journal of General Psychology* 137.4, pp. 362–375. ISSN: 00221309. DOI: 10.1080/00221309.2010.499397.

Rosenfeld, Avi et al. (2014). 'Learning Driver's Behavior to Improve Adaptive Cruise Control'. In: *Journal of Intelligent Transportation Systems* 2450.July 2015, pp. 18–31. ISSN: 1547-2450. DOI: 10.1080/15472450.2014.889960.

Rosenthal, Robert (1976). *Experimenter effects in behavioral research.* Enlarged e. Oxford, England: Irvington.

Roth, Robert E. (2017). 'Visual Variables'. In: *The International Encyclopedia of Geography.* Ed. by Douglas Richardson et al. John Wiley and Sohns, Ltd. ISBN: 9781118786352. DOI: 10.1002/9781118786352.wbieg0761.

Rothgeb, Jennifer M. (2008). 'Pilot Test'. In: *Encyclopedia of Survey Research Methods.* Ed. by Paul J. Lavrakas. Thousand Oaks: SAGE Publications, Inc., pp. 584–585. ISBN: 9781412963947. DOI: 10.4135/9781412963947.n377.

Sadeghian Borojeni, Shadan et al. (2018). 'Feel the Movement: Real Motion Influences Responses to Take-over Requests in Highly Automated Vehicles'. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* 246, pp. 1–13. DOI: 10.1145/3173574.3173820.

SAE International (2014). *SAE International's Levels of Driving Automation for on-road vehicles (J3016).* URL: http://www.sae.org/misc/pdfs/automated_driving.pdf. (visited on 07/11/2016).

– (2015). *Surface Vehicle Recommended Practice: Operational Definitions of Driving Performance Measures and Statistics (J2944_201506).* URL: https://saemobilus.sae.org/content/j2944_201506 (visited on 26/10/2019).

– (2018). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016_201806).* URL: https://www.sae.org/standards/content/j3016_201806/ (visited on 01/07/2018).

Saffarian, M., Joost C.F. de Winter and R. Happee (2012). 'Automated Driving: Human-Factors Issues and Design Solutions'. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56.1, pp. 2296–2300. ISSN: 1071-1813. DOI: 10.1177/1071181312561483.

Salas, Eduardo et al. (1995). 'Situation Awareness in Team Performance: Implications for Measurement and Training'. In: *Human Factors* 37.1, pp. 123–136. ISSN: 00187208. DOI: 10.1518/001872095779049525.

Salmon, Paul M., Neville A. Stanton, Guy H. Walker, Daniel Jenkins et al. (2009). 'Measuring Situation Awareness in complex systems: Comparison of measures study'. In: *International*

Journal of Industrial Ergonomics 39.3, pp. 490–500. ISSN: 01698141. DOI: 10.1016/j.ergon.2008.10.010.

Salmon, Paul M., Neville A. Stanton, Guy Walker and Damian Green (2006). 'Situation awareness measurement: A review of applicability for C4i environments'. In: *Applied Ergonomics* 37.2, pp. 225–238. ISSN: 00036870. DOI: 10.1016/j.apergo.2005.02.001.

Sardegna, Jill et al. (2002). *The Encyclopedia of Blindness and Vision Impairment*. 2nd ed. New York, NY, US: Facts On File. ISBN: 978-0816042807.

Sarter, Nadine B., Ann Arbor et al. (2007). 'Pilots' Monitoring Strategies and Performance on Automated Flight Decks: An Empirical Study Combining Behavioral and Eye-Tracking Data'. In: *Human Factors* 49.3, pp. 347–357. DOI: 10.1518/001872007X196685.

Sarter, Nadine B. and David D. Woods (1995). 'How in the World Did We Ever Get into That Mode? Mode Error and Awareness in Supervisory Control'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37.1, pp. 5–19. ISSN: 00187208. DOI: 10.1518/001872095779049516.

Schaefer, Kristin E. et al. (2016). 'A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 58.3, pp. 377–400. ISSN: 0018-7208. DOI: 10.1177/0018720816634228.

Schaller, Thomas, Joachim Schiehlen and Barbara Gradenegger (2008). 'Congestion Assistance - Supporting the Driver in Lateral and Longitudinal Guidance. System Development and Customer Acceptance'. In: *Aktive Sicherheit durch Fahrerassistenz*. Muenchen.

Scherer, Klaus R. (2005). 'What are emotions? And how can they be measured?' In: *Social Science Information* 44.4, pp. 695–729. DOI: 10.1177/0539018405058216.

Schwalk, Matti, Niko Kalogerakis and Thomas Maier (2015). 'Driver Support by a Vibrotactile Seat Matrix - Recognition, Adequacy and Workload of Tactile Patterns in Take-over Scenarios During Automated Driving'. In: *Procedia Manufacturing* 3, pp. 2466–2473. ISSN: 23519789. DOI: 10.1016/j.promfg.2015.07.507.

Schweber, Bill (2016). *The Autonomous Car: A Diverse Array of Sensors Drives Navigation, Driving, and Performance*. URL: http://www.mouser.co.uk/applications/autonomous-car-sensors-drive-performance/ (visited on 05/03/2017).

Seffah, Ahmed et al. (2006). 'Usability measurement and metrics: A consolidated model'. In: *Software Quality Journal* 14.2, pp. 159–178. ISSN: 15731367. DOI: 10.1007/s11219-006-7600-8.

Self, Brian P. et al. (2008). 'Human factors issues of tactile displays for military environments'. In: *Tactile Displays for Orientation, Navigation and Communication in Air, Sea and Land Environments* January, pp. 1–18.

Seppelt, Bobbie D. and John D. Lee (2007). 'Making adaptive cruise control (ACC) limits visible'. In: *International Journal of Human Computer Studies* 65.3, pp. 192–205. ISSN: 10715819. DOI: 10.1016/j.ijhcs.2006.10.001.

Seppelt, Bobbie D., Sean Seaman et al. (2017). 'Glass half-full: On-road glance metrics differentiate crashes from near-crashes in the 100-Car data'. In: *Accident Analysis and Prevention* 107.December 2016, pp. 48–62. ISSN: 00014575. DOI: 10.1016/j.aap.2017.07.021.

Shaffer, Fred and J. P. Ginsberg (2017). 'An Overview of Heart Rate Variability Metrics and Norms'. In: *Frontiers in Public Health* 5.September, pp. 1–17. ISSN: 2296-2565. DOI: 10.3389/fpubh.2017.00258.

Shami, N. Sadat, Gilly Lshed and David Klein (2005). 'Context of Use Evaluation of Peripheral Displays'. In: *Proceedings of the IFIP TC13 International Conference on Human Computer In- teraction (INTERACT)*. Ed. by Maria Francesca Costabile and Fabio Paterno. New York: Springer, pp. 579–587. ISBN: 9783642130939. DOI: 10.1007/3-540-68339-9_34.

Shapiro, Gary M. (2008a). 'Sample'. In: *Encyclopedia of Survey Research Methods*. Ed. by Paul J. Lavrakas. Thousand Oaks: SAGE Publications, Inc., pp. 776–777. ISBN: 9781412918084. DOI: 10.4135/9781412963947.n502.

– (2008b). 'Sample Design'. In: *Encyclopedia of Survey Research Methods*. Ed. by Paul J. Lavrakas. Thousand Oaks: SAGE Publications, Inc., p. 777. ISBN: 9781412963947. DOI: 10.4135/9781412963947.n503.

Shepardson, David (2018). *Tesla hits parked California police vehicle; driver blames 'Autopilot'*. URL: https://www.reuters.com/article/us-tesla-autopilot/tesla-hits-parked-california-police-vehicle-driver-blames-autopilot-idUSKCN1IU2SZ (visited on 30/05/2018).

Sheridan, Thomas B. (2002). *Humans and automation: system design and research issues*. Santa Monica, CA: John Wiley, in cooperation with the Human Factors and Ergonomics Society. ISBN: 0471234281.

Shinohara, Kazumitsu et al. (2010). 'Detailed analysis of distraction induced by in-vehicle verbal interactions on visual search performance'. In: *IATSS Research* 34.1, pp. 42–47. ISSN: 08885885. DOI: 10.1021/ie301168r.

Simons, Daniel J. and Christopher F. Chabris (1999). 'Gorillas in our midst: sustained inattentional blindness for dynamic events'. In: *Perception* 28, pp. 1059–1074. DOI: 10.1068/p281059.

Simpson, Michael J. (2017). 'Mini-review: Far peripheral vision'. In: *Vision Research* 140, pp. 96–105. ISSN: 18785646. DOI: 10.1016/j.visres.2017.08.001.

Singh, Indramani L., Robert Molloy and Raja Parasuraman (1993). 'Automation-Induced "Complacency": Development of the Complacency-Potential Rating Scale'. In: *The International Journal of Aviation Psychology* 3.2, pp. 111–122. DOI: 10.1207/s15327108ijap0302_2.

Snijders, Tom A. B. and Roel J. Bosker (2012). *Multilevel Analysis*. London: SAGE Publications Ltd, p. 368. ISBN: 978-1849202015.

Stanton, Neville A., Alain Dunoyer and Adam Leatherland (2011). 'Detection of new in-path targets by drivers using Stop & Go Adaptive Cruise Control'. In: *Applied Ergonomics* 42.4, pp. 592–601. ISSN: 00036870. DOI: 10.1016/j.apergo.2010.08.016.

Stanton, Neville A., Paul M. Salmon, Guy H. Walker and Daniel P. Jenkins (2010). 'Is situation awareness all in the mind?' In: *Theoretical Issues in Ergonomics Science* 11.1-2, pp. 29–40. ISSN: 1463-922X. DOI: 10.1080/14639220903009938.

Stanton, Neville A., Paul M. Salmon, Guy H. Walker, Eduardo Salas et al. (2017). 'State-of-Science: Situation Awareness in individuals, teams and systems'. In: *Ergonomics* 0139.February, pp. 1–33. ISSN: 0014-0139. DOI: 10.1080/00140139.2017.1278796. URL: https://www.tandfonline.com/doi/full/10.1080/00140139.2017.1278796.

Stanton, Neville A., Rebecca Stewart et al. (2006). 'Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology'. In: *Ergonomics* 49.12-13, pp. 1288–1311. ISSN: 1288–1311 Distributed. DOI: 10.1080/00140130600612762.

Stasko, John et al. (2004). 'Personalized Peripheral Information Awareness Through Information Art'. In: *UbiComp 2004: Ubiquitous Computing. UbiComp 2004. Lecture Notes in Computer Science*. Berlin: Springer, pp. 18–35. ISBN: 978-3-540-30119-6. DOI: https://doi.org/10.1007/978-3-540-30119-6_2.

Steptoe, A. and C. Voegele (1991). 'Methodology of mental stress testing in cardiovascular research'. In: *Circulation* 83.4 Suppl, pp. 14–24.

Stewart, Jack (2018). *Why Tesla's Autopilot Can't See a Stopped Firetruck*. URL: https://www.wired.com/story/tesla-autopilot-why-crash-radar (visited on 23/05/2018).

Strand, Niklas et al. (2014). 'Semi-automated versus highly automated driving in critical situations caused by automation failures'. In: *Transportation Research Part F: Traffic Psychology and Behaviour* 27.PB, pp. 218–228. ISSN: 13698478. DOI: 10.1016/j.trf.2014.04.005.

Strasburger, Hans, Ingo Rentschler and Martin Jüttner (2011). 'Peripheral vision and pattern recognition: A review'. In: *Journal of vision* 11.5, p. 13. ISSN: 1534-7362. DOI: 10.1167/11.5.13.

Summerskill, Stephen J. et al. (2019). *The definition, production and validation of the direct vision standard (DVS) for HGVS. Final Report for TfL review. Version 1.1.* Tech. rep. London: Transport for London.

Systems Technology, Inc (2017). *STISIM Drive 3 [Computer software].*

Tabachnick, Barbara G. and Linda S. Fidell (2013). *Using Multivariate Statistics*. 6th ed. [S.l.]: Pearson. ISBN: 9780205849574.

Tang, Kuo Hao and Yueh Hua Lee (2016). 'Evaluation of detection and discrimination ability of peripheral vision on notification information based on large displays'. In: *Displays* 41, pp. 50–60. ISSN: 01419382. DOI: 10.1016/j.displa.2015.12.002.

Taylor, R.M. (1990). 'Situational awareness rating technique (SART): the development of a tool for aircrew systems design'. In: *Situational Awareness in Aerospace Operations (AGARD-CP- 478).* Neuilly Sur Seine, France, pp. 3/1–3/17.

Telpaz, Ariel et al. (2015). 'Haptic Seat for Automated Driving : Preparing the Driver to Take Control Effectively'. In: *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 23–30. DOI: 10.1145/2799250.2799267.

Thomson, Judi et al. (2005). 'A Typology for Visualizing Uncertainty'. In: *Visualization and Data Analysis* 5669.January, pp. 146–157. ISSN: 0277786X. DOI: 10.1117/12.587254.

Thüring, Manfred and Sascha Mahlke (2007). 'Usability, aesthetics and emotions in human-technology interaction'. In: *International Journal of Psychology* 42.4, pp. 253–264. ISSN: 00207594. DOI: 10.1080/00207590701396674.

Tobii AB (2017). *Tobii Pro Lab User Manual v.1.79.* URL: https://www.tobiipro.com/learn-and-support/downloads-pro/ (visited on 17/10/2017).

– (2018a). *How do Tobii Eye Trackers work?* URL: https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/how-do-tobii-eye-trackers-work/ (visited on 24/07/2019).

– (2018b). *Tobii Lab Pro (Version 1.86) [Computer software].*

– (2019a). *What does eye tracking data tell us?* URL: https://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/interpret/what-does-eye-tracking-data-tell-us/ (visited on 12/07/2019).

– (2019b). *What is eye-tracking?* URL: https://www.tobiipro.com/blog/what-is-eye-tracking/ (visited on 12/07/2019).

Topliss, Bethan Hannah et al. (2018). 'Establishing the Role of a Virtual Lead Vehicle as a Novel Augmented Reality Navigational Aid'. In: *Proceedings of the 10th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '18).* Toronto, ON, Canada: ACM, pp. 137–145. DOI: 10.1145/3239060.3239069.

Treisman, Anne M. and Garry Gelade (1980). 'A Feature-Integration Theory of Attention'. In: *Cognitive Psychology* 12.1, pp. 97–136. ISSN: 00100285. DOI: 10.1016/0010-0285(80)90005-5.

Treisman, Anne M. and Stephen Gormican (1988). 'Feature analysis in early vision - evidence from search asymmetries..pdf'. In: *Psychological review* 95.1, pp. 15–48. ISSN: 0033-295X. DOI: 10.1037/0033-295X.95.1.15.

Trimble, T. E. et al. (2014). *Human Factors Evaluation of Level 2 and Level 3 Automated Driving Concepts: Past Research, State of Automation Technology, and Emerging System Concepts*. Tech. rep. July. Washington, DC: National Highway Traffic Safety Administration.

Trobia, Alberto (2008). 'Encyclopedia of Survey Research Methods: Cronbach's Alpha'. In: *SAGE*, pp. 169–170. DOI: http://dx.doi.org/10.4135/9781412963947.n117.

Trösterer, Sandra, Magdalena Gärtner et al. (2016). 'You Never Forget How to Drive: Driver Skilling and Deskilling in the Advent of Autonomous Vehicles'. In: *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. Ann Arbor, MI, USA: ACM, pp. 209–216. ISBN: 9781450345330.

Trösterer, Sandra, Alexander Meschtscherjakov et al. (2017). 'What We Can Learn from Pilots for Handovers and (De)Skilling in Semi-Autonomous Driving'. In: *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '17*, pp. 173–182. DOI: 10.1145/3122986.3123020.

Uggirala, Ananth et al. (2004). 'Measurement of trust in complex and dynamic systems using a quantitative approach'. In: *International Journal of Industrial Ergonomics* 34, pp. 175–186. DOI: 10.1016/j.ergon.2004.03.005.

UNECE (2017). *Discussion paper on possible driver's "other activities" while an automated driving system is engaged.* URL: https://www.unece.org/fileadmin/DAM/trans/doc/2017/wp1/ECE-TRANS-WP1-S-INF-2017-4e.pdf (visited on 26/10/2019).

– (2019). *Proposal for amendments to ECE/TRANS/WP.29/2019/34. Framework document on automated/autonomous vehicles (levels 3 and higher).* URL: http://www.unece.org/fileadmin/DAM/trans/doc/2019/wp29/WP.29-178-10r2e.pdf (visited on 26/10/2019).

Van Den Beukel, Arie P. and Mascha C. Van Der Voort (2014). 'Design Considerations on User-Interaction for Semi-Automated Driving'. In: *FISITA 2014 World Automotive Congress*, pp. 1–8.

Van Erp, Jan B. F. and Hendrik A. H. C. Van Veen (2001). 'Vibro-tactile information presentation in automobiles'. In: *Proceedings of Eurohaptics*, pp. 99–104.

Vatavu, Radu Daniel (2013). 'On designing interactivity awareness for ambient displays'. In: *Multimedia Tools and Applications* 66.1, pp. 59–80. ISSN: 13807501. DOI: 10.1007/s11042-012-1140-y.

Visser, Ewart J. de, Frank Krueger et al. (2012). 'The world is not enough: Trust in cognitive agents'. In: *Proceedings of the Human Factors and Ergonomics Society*, pp. 263–267. ISSN: 10711813. DOI: 10.1177/1071181312561062.

Visser, Ewart J. de, Samuel S. Monfort et al. (2017). 'A Little Anthropomorphism Goes a Long Way: Effects of Oxytocin on Trust, Compliance, and Team Performance With Automated Agents'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 59.1, pp. 116–133. ISSN: 1940-8153. DOI: 10.1177/0018720816687205.

Visser, Ewart J. de, Richard Pak and Tyler H. Shaw (2018). 'From 'automation' to 'autonomy': the importance of trust repair in human–machine interaction'. In: *Ergonomics* 0139, pp. 1–19. ISSN: 13665847. DOI: 10.1080/00140139.2018.1457725.

Vogel, Katja (2003). 'A comparison of headway and time to collision as safety indicators'. In: *Accident Analysis and Prevention* 35.3, pp. 427–433. ISSN: 00014575. DOI: 10.1016/S0001-4575(02)00022-2.

Vollrath, Mark, Susanne Schleicher and Christhard Gelau (2011). 'The influence of Cruise Control and Adaptive Cruise Control on driving behaviour - A driving simulator study'. In: *Accident Analysis and Prevention* 43.3, pp. 1134–1139. ISSN: 00014575. DOI: 10.1016/j.aap.2010.12.023.

Vredenburg, Karel et al. (2002). 'A survey of user-centered design practice'. In: *Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world, changing ourselves - CHI '02* 1, p. 471. ISSN: 10629432. DOI: 10.1145/503376.503460.

Vrtunski, P. Bart, Marian B. Patterson and Gregory O. Hill (1984). 'Factor Analysis of Choice Reaction Time in Young and Elderly Subjects'. In: *Perceptual and Motor Skills* 59, pp. 659–676.

Waard, Dick de et al. (1999). 'Driver Behavior in an Emergency Situation in the Automated Highway System'. In: *Transportation Human Factors* 1.1, pp. 67–82. ISSN: 1093-9741. DOI: 10.1207/sthf0101_9.

Walch, Marcel et al. (2016). 'Towards Cooperative Driving: Involving the Driver in an Autonomous Vehicle's Decision Making'. In: *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. November. Ann Arbor, MI, USA: ACM. ISBN: 9781450345330. DOI: 10.1145/3003715.3005458.

Wandtner, Bernhard, Nadja Schömig and Gerald Schmidt (2018). 'Effects of Non-Driving Related Task Modalities on Takeover Performance in Highly Automated Driving'. In: *Human Factors* 60.6, pp. 870–881. ISSN: 15478181. DOI: 10.1177/0018720818768199.

Wang, Lu, Greg A. Jamieson and Justin G. Hollands (2009). 'Trust and reliance on an automated combat identification system'. In: *Human factors* 51.3, pp. 281–291. ISSN: 0018-7208. DOI: 10.1177/0018720809338842.

Wang, X., R. A. Pinol et al. (2014). 'Optogenetic Stimulation of Locus Ceruleus Neurons Augments Inhibitory Transmission to Parasympathetic Cardiac Vagal Neurons via Activation of Brainstem 1 and 1 Receptors'. In: *Journal of Neuroscience* 34.18, pp. 6182–6189. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.5093-13.2014.

Ware, Colin (2012). *Information Visualization: Perception for Design*. 3rd ed. Elsevier Science & Technology. ISBN: 9780123814654.

– (2013). 'Visual Salience and Finding Information'. In: *Information Visualization*, pp. 139–177. ISBN: 978-0-12-381464-7. DOI: 10.1016/B978-0-12-381464-7.00005-3.

Waytz, Adam, Joy Heafner and Nicholas Epley (2014). 'The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle'. In: *Journal of Experimental Social Psychology* 52, pp. 113–117. ISSN: 00221031. DOI: 10.1016/j.jesp.2014.01.005.

Weiser, Mark and John Seely Brown (1996). 'Designing Calm Technology'. In: *PowerGrid Journal* 1.1, pp. 75–85. ISSN: 0735-6331. DOI: 10.1.1.135.9788.

West, Karl (2019). *Dieselgate leaves UK's car industry in crisis*. URL: https://www.theguardian.com/business/2018/apr/21/dieselgate-uk-car-industry-sales-slump (visited on 01/04/2019).

Wickens, Christopher D. (2002). 'Multiple resources and performance prediction'. In: *Theoretical Issues in Ergonomics Science* 3.2, pp. 159–177. ISSN: 1463-922X. DOI: 10.1080/14639220210123806.

– (2005). 'Attentional Tunneling and Task Management'. In: *Technical Report AHFD-05-01/NASA-05-10* December, pp. 812–817.

– (2008). 'Multiple Resources and Mental Workload'. In: *Human Factors* 50.3, pp. 449–455. ISSN: 0018-7208. DOI: 10.1518/001872008X288394..

Wickens, Christopher D. and C. Melody Carswell (1995). 'The Proximity Compatibility Principle: Its Psychological Foundation and Relevance to Display Design'. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37.3, pp. 473–494. ISSN: 1547-8181. DOI: 10.1518/001872095779049408.

– (2012). 'Information Processing'. In: *Handbook of Human Factors and Ergonomics*. 4th ed. New York, NY, US: Wiley. Chap. 5. DOI: 10.1002/9781118131350.ch5.

Wickens, Christopher D. and Stephen R. Dixon (2007). 'The benefits of imperfect diagnostic automation: a synthesis of the literature'. In: *Theoretical Issues in Ergonomics Science* 8.3, pp. 201–212. ISSN: 1463-922X. DOI: 10.1080/14639220500370105.

Wickens, Christopher D. and Justin G. Hollands (2000). *Engineering Psychology and Human Performance*. 3rd ed. Upper Saddle River, NJ: Prentice Hall, p. 573. ISBN: 9780321047113.

Wickens, Christopher D., Justin G. Hollands et al. (2013). *Engineering Psychology and Human Performance.* 4. ed., in. Boston, MA: Pearson Education, Inc. ISBN: 9780205945740.

Wickens, Christopher D. and Colin Kessel (1979). 'The Effects of Participatory Mode and Task Workload on the Detection of Dynamic System Failures'. In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1, pp. 24–34. ISSN: 0018-9472. DOI: 10.1109/TSMC.1979.4310070.

Wickens, Christopher D. and Jason S. McCarley (2008). *Applied Attention Theory.* Boca Raton, Florida: CRC Press, p. 248. ISBN: 9780805859836.

Wickens, Christopher D., Angelia Sebok et al. (2015). 'Using Modeling and Simulation to Predict Operator Performance and Automation-Induced Complacency with Robotic Automation'. In: *Human Factors* 57.6, pp. 959–975. ISSN: 15478181. DOI: 10.1177/0018720814566454.

Wiener, Earl L. (1989). *Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft.* Miami. URL: http://ntrs.nasa.gov/search.jsp?R=19890016609 (visited on 29/03/2019).

Wiener, Earl L. and Renwick E. Curry (1980). 'Flight-deck automation: promises and problems'. In: *Ergonomics* 23.10, pp. 995–1011. DOI: 10.1080/00140138008924809.

Wierwille, Walter W. (June 1993). 'Demands on driver resources associated with introducing advanced technology into the vehicle'. In: *Transportation Research Part C: Emerging Technologies* 1.2, pp. 133–142. ISSN: 0968090X. DOI: 10.1016/0968-090X(93)90010-D.

Wilde, Gerald J. S. (1998). 'Risk homeostasis theory: an overview.' In: *Injury prevention: Journal of the International Society for Child and Adolescent Injury Prevention* 4.2, pp. 89–91. ISSN: 1353-8047. DOI: 10.1136/ip.4.2.89.

Winer, Ben J. (1971). *Statistical principles in experimental design.* New York, NY: McGraw-Hill.

Winkler, Susann, Juela Kazazi and Mark Vollrath (2018). 'How to warn drivers in various safety-critical situations - Different strategies, different reactions'. In: *Accident Analysis and Prevention* 117. DOI: 10.1016/j.aap.2018.01.040.

Winner, Hermann and Stephan Hakuli (2012). *Handbuch Fahrerassistenzsysteme.* Ed. by Gabriele Wolf. 2nd. Wiesbaden: Vieweg+Teubner Verlag. ISBN: 9783834814579. DOI: 10.1007/978-3-8348-8619-4.

Wintersberger, Philipp, Dmitrijs Dmitrenko et al. (2019). 'S(C)Entinel'. In: *Proceedings of the 24th International Conference on Intelligent User Interfaces - IUI '19* March, pp. 538–546. DOI: 10.1145/3301275.3302332.

Wintersberger, Philipp, Tamara von Sawitzky et al. (2017). 'Traffic Augmentation as a Means to Increase Trust in Automated Driving Systems'. In: *Proceedings of the 12th*

Biannual Conference on Italian SIGCHI Chapter - CHItaly '17 September, pp. 1–7. DOI: 10.1145/3125571.3125600.

Wisneski, Craig et al. (1998). 'Ambient Displays: Turning Architectural Space into an Interface between People and Digital Information'. In: *Cooperative Buildings*. Ed. by Gerhard Goos, Juris Hartmanis and Jan van Leeuwen. Berlin: Springer, pp. 22–32. ISBN: 9783642130939. DOI: 10.1007/3-540-68339-9_34.

Woods, David D. (1993). 'The Price of Flexibility'. In: *IUI '93 Proceedings of the 1st international conference on Intelligent user interfaces*. Orlando, Florida: ACM New York, pp. 19–25. DOI: 10.1145/169891.169894.

– (2016). 'The Risks of Autonomy: Doyles Catch'. In: *Journal of Cognitive Engineering and Decision Making* 10.2, pp. 131–133. ISSN: 1555-3434. DOI: 10.1177/1555343416653562.

Woods, David D. and Richard I. Cook (2006). 'Incidents - Markers of Resilience or Brittleness?' In: *Resilience Engineering: Concepts and Precepts*. Boca Raton, Florida: CRC Press, pp. 61–67. ISBN: 978-0754646419.

Woolrych, Alan and Gilbert Cockton (2001). 'Why and when five test users aren't enough'. In: *Proceedings of IHM-HCI 2001 conference* 2.January, pp. 105–108.

World Health Organization (2018). 'Global Status Report on Road Safety'. In: *Licence: CC BY-NC-SA 3.0 IGO*. Geneva: World Health Organization.

Yang, X. Jessie et al. (2017). 'Evaluating Effects of User Experience and System Transparency on Trust in Automation'. In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17*, pp. 408–416. ISSN: 21672148. DOI: 10.1145/2909824.3020230.

Yerkes, Robert M. and John D. Dodson (Nov. 1908). 'The relation of strength of stimulus to rapidity of habit-formation'. In: *Journal of Comparative Neurology and Psychology* 18.5, pp. 459–482. ISSN: 0092-7015. DOI: 10.1002/cne.920180503.

Yoshida, Junko (2016). *Deep Learning: Achilles Heel in Robo-Car Tests*. (Visited on 27/02/2017).

Young, L. R. (1969). 'On Adaptive Manual Control'. In: *IEEE Transactions on Man-Machine Systems* 10.4, pp. 292–331. ISSN: 21682860. DOI: 10.1109/TMMS.1969.299931.

Young, Mark S. and Neville A. Stanton (2002). 'Attention and automation: New perspectives on mental underload and performance'. In: *Theoretical Issues in Ergonomics Science* 3.2, pp. 178–194. ISSN: 1464536X. DOI: 10.1080/14639220210123789.

– (2007). 'Back to the future: brake reaction times for manual and automated vehicles.' In: *Ergonomics* 50.1, pp. 46–58. ISSN: 0014-0139. DOI: 10.1080/00140130600980789.

Young, Rob (2019). *Ehiopian Airlines crash: 'Pitch up, pitch up!"* URL: https://www.bbc.com/news/business-47759966 (visited on 31/03/2019).

Zaman, Kais and Sankaran Mahadevan (2017). 'Reliability-based design optimization of multidisciplinary system under aleatory and epistemic uncertainty'. In: *Structural and Multidisciplinary Optimization* 55.2, pp. 681–699. ISSN: 16151488. DOI: 10.1007/s00158-016-1532-0.

Zeeb, Kathrin, Axel Buchner and Michael Schrauf (2015). 'What determines the take-over time? An integrated model approach of driver take-over after automated driving'. In: *Accident Analysis and Prevention* 78, pp. 212–221. ISSN: 00014575. DOI: 10.1016/j.aap.2015.02.023.

Glossary

This glossary provides explanations of terms relevant to this thesis. The definitions are to be regarded as direct quotes from the referenced publications.

A | C | D | E | F | G | H | L | M | O | R | S | T | U | V

A

Accuracy

The extent to which an actual outcome matches an intended outcome (ISO 2018).

Active safety system

Active safety systems are vehicle systems that sense and monitor conditions inside and outside the vehicle for the purpose of identifying perceived present and potential dangers to the vehicle, occupants, and/or other road users, and automatically intervene to help avoid or mitigate potential collisions via various methods, including alerts to the driver, vehicle system adjustments, and/or active control of the vehicle subsystems (brake, throttle, suspension, etc.) (SAE International 2018).

Adaptive cruise control

Technology that allows a vehicle to automatically adjust its speed to maintain a preset distance from the vehicle in front of it based on the driver's preferences (Rosenfeld et al. 2014).

ADS-dedicated vehicle

A vehicle designed to be operated exclusively by a level 4 or level 5 automated driving system (ADS) for all trips. An ADS-dedicated vehicle (ADS-DV) might be designed without user interfaces, such as braking, accelerating, steering, and transmission gear selection input devices designed to be operable by a human driver (SAE International 2018).

Area of interest

Physical location where specific task-related information can be found (Wickens, Hollands et al. 2013, p. 50).

Attention

In the context of cognitive psychology, attention serves as a filter of information when selecting which sensations are to be further processed and supplies all human information processing stages with mental resources (Wickens and McCarley 2008).

Automated driving system

The hardware and software that are collectively capable of performing the entire DDT on a sustained basis, regardless of whether it is limited to a specific ODD; this term is used specifically to describe an SAE level 3, 4, or 5 driving automation system (SAE International 2018).

Automation bias

The tendency of human operators to place greater faith in the automation than in other information sources. Consequently, users do not respond to critical situations if not alerted (error of omission) or follow an erratic advice (commission error) (Parasuraman and Manzey 2010).

C

Complacency

An active reallocation of attention away from the automation to other tasks [...] under conditions of multiple-task load, when manual tasks compete with the automated task for the operator's attention (Parasuraman and Manzey 2010, p. 384).

Context of use

Combination of users, goals and tasks, resources, and environment (ISO 2018).

Cry wolf effect

Repeated false alarms as a consequence of a low threshold to trigger an alarm that can lead to the disuse of automation (Yang et al. 2017).

D

DDT fallback

The response by the user or by an ADS to either perform the DDT or achieve a minimal risk condition after the occurrence of a DDT performance-relevant system failure or upon DDT exit (SAE International 2018).

DDT fallback-ready user

The user of a vehicle equipped with an engaged level 3 ADS feature who is able to operate the vehicle and is receptive to ADS-issued requests to intervene and to evident DDT performance-relevant system failures in the vehicle compelling him or her to perform the DDT fallback (SAE International 2018).

Display

Superordinate designation for technical elements that convey information from the machine to the user, i. e. the counterparts of controls.

Driving automation

The performance of part or all of the DDT on a sustained basis (SAE International 2018).

Driving automation system

The hardware and software that are collectively capable of performing part or all of the DDT on a sustained basis; this term is used generically to describe any system capable of level 1-5 driving automation (SAE International 2018).

Driving mode

A type of vehicle operation with characteristic DDT requirements (e. g., expressway merging, high-speed cruising, low-speed traffic jam, etc.) (SAE International 2018).

Dynamic driving task

Subset of the driving task that includes operational (steering, braking, accelerating, monitoring the vehicle and roadway) and tactical (responding to events, determining when to change lanes, turn, use signals, etc.) aspects but not the strategical (determining destinations and waypoints) component (SAE International 2014).

E

Effectiveness

Accuracy and completeness with which users achieve specified goals (ISO 2018).

Efficiency

Resources used in relation to the results achieved (ISO 2018).

F

Feature

A driving automation system's design-specific functionality at a specific level of driving automation within a particular ODD. A given driving automation system may have multiple features, each associated with a particular level of driving automation and ODD (SAE International 2018).

Function-specific uncertainty

Within the context of this thesis, function-specific uncertainty delineates the reliability of individual DDT functions.

G

Goal

Intended outcome (ISO 2018).

Ground truth

In the context of SA: referring to the objective state of the world or the objective unfolding of events that are predicted against which the accuracy of SA can be assessed (Parasuraman, Sheridan and Wickens 2008).

Guessability

A measure of the cost to the user in using a product to perform a new task for the first time – the lower the cost (for example, in terms of time on task or errors made) the higher the guessability (Jordan 1998, p. 11).

H

Human-machine interface

Any points of interaction between human and machine where information is transferred from the human operator to the machine or vice versa (Bubb et al. 2015). Human operators accomplish this information transfer via controls that generate an input to the machine. Machines convey information by producing an output which is perceivable by humans. This thesis uses the general term displays as the superordinate designation for technical elements that convey information from the machine to the user, i. e. the counterparts of controls.

L

Lateral vehicle motion

The DDT subtask comprising the activities necessary for the real-time, sustained regulation of the y-axis component of vehicle motion. Lateral vehicle motion control includes the detection of the vehicle positioning relative to lane boundaries and application of steering and/or differential braking inputs to maintain appropriate lateral positioning (SAE International 2018).

Longitudinal vehicle motion

The DDT subtask comprising the activities necessary for the real-time, sustained regulation of the x-axis component of vehicle motion. Longitudinal vehicle control includes maintaining a set speed as well as detecting a preceding vehicle in the path of the subject vehicle, maintaining an appropriate gap to the preceding vehicle, and applying propulsion or braking inputs to cause the vehicle to maintain that speed or gap (SAE International 2018).

M

Mental model

A mental structure that reflects the user's understanding of a system and therefore is a source of expectancies about how a system will respond (Carroll and Olson 1987, p. 12; as cited in Wickens, Hollands et al. 2013, p. 236).

Mental workload

Mental workload characterises the demands of tasks imposed on the limited information processing capacity of the brain (Wickens, Hollands et al. 2013, p. 347).

Minimal risk condition

A condition to which a user or an ADS may bring a vehicle after performing the DDT fallback in order to reduce the risk of a crash when a given trip cannot or should not be completed (SAE International 2018).

Mode awareness

The ability of a supervisor to track and to anticipate the behaviour of automated systems (Sarter and Woods 1995, p. 6).

Mode error

A human-machine system breakdown, that requires that the users lose track of which mode the device is in (or confuse which methods or actions are appropriate for which mode) and requires a machine for which the same actions and indications mean different things in different modes of operation (Sarter and Woods 1995, p. 6).

Monitor

A general term referencing a range of functions involving real-time human or machine sensing and processing of data used to operate a vehicle, or to support its operation (SAE International 2018).

Monitor driving automation system performance

The activities and/or automated routines for evaluating whether the driving automation system is performing part or all of the DDT appropriately (SAE International 2018).

Monitor the driving environment

The activities and/or automated routines that accomplish real-time roadway environmental object and event detection, recognition, classification, and response preparation (excluding actual response), as needed to operate a vehicle (SAE International 2018).

Monitor the user

The activities and/or automated routines designed to assess whether and to what degree the user is performing the role specified for him/her (SAE International 2018).

Monitor vehicle performance

The activities and/or automated routines that accomplish real-time evaluation of the vehicle performance, and response preparation, as needed to operate a vehicle (SAE International 2018).

O

Object and event detection and response

The subtasks of the DDT that include monitoring the driving environment (detecting, recognising, and classifying objects and events and preparing to respond as needed) and executing an appropriate response to such objects and events (i. e. as needed to complete the DDT and/or DDT fallback) (SAE International 2018).

OOTL performance problem

A major potential consequence of automation that leaves operators of automated systems handicapped in their ability to take over manual operations in the event of automation failure (Endsley and Kiris 1995, p. 381).

Operate

Collectively, the activities performed by a (human) driver (with or without support from one or more level 1 or 2 driving automation features) or by an ADS (level 3-5) to perform the entire DDT for a given vehicle during a trip (SAE International 2018).

Operational design domain

The specific conditions under which a given driving system or feature thereof is designed to function, including, but not limited to, driving modes. An ODD may include geographic, roadway, environmental, traffic, speed, and/or temporal limitations (SAE International 2018).

R

Receptivity

An aspect of consciousness characterised by a person's ability to reliably and appropriately focus his/her attention in response to a stimulus. In level 0-2 driving automation, the driver is expected to be receptive to evident vehicle system failures, such as a broken tie rod. In level 3 driving automation, a DDT fallback-ready user is considered to be receptive to a request to intervene and/or to an evident vehicle system failure, whether or not the ADS issues a request to intervene as a result of such a vehicle system failure (SAE International 2018).

Request to intervene (TOR)

Notification by the automated driving system to a human driver that he/she should promptly begin or resume performance of the DDT (SAE International 2014).

Resource

The context of use includes any reusable resources and any expendable and/or exhaustible resources that need to be used to achieve specified goals (ISO 2018).

Risk homeostasis

The willingness to take a certain amount of risk and related behavioural adaptations to changes in environmental conditions (Wilde 1998).

S

Satisfaction

Extent to which the user's physical, cognitive, and emotional responses that result from the use of a system, product, or service meet the user's needs and expectations (ISO 2018).

Situation awareness

The perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future (Endsley 1988a, p. 97).

Supervise

The driver activities, performed while operating a vehicle with an engaged level 1 or 2 driving automation system, to monitor the driving automation system's performance, respond to inappropriate actions taken by that system, and to otherwise complete the DDT (SAE International 2018).

Sustained

Performance of part or all of the DDT both between and across external events, including responding to external events and continuing performance of part or all of the DDT in the absence of external events (SAE International 2018).

System

Refers to the driver assistance system, combination of driver assistance systems, or automated driving system. Excluded are warning and momentary intervention systems, which do not automate any part of the dynamic driving task on a sustained basis and therefore do not change the human driver's role in performing the dynamic driving task (SAE International 2014).

System failure

A malfunction in a driving automation system and/or other vehicle system that prevents the driving automation system from reliably sustaining (partial or complete) DDT performance (SAE International 2018).

System-wide uncertainty

Within the context of this thesis, system-wide uncertainty is defined as a time-specific indicator for the overall reliability of the automated driving system.

T

Task

Set of activities undertaken in order to achieve a specific goal (ISO 2018).

Trip

The traversal of an entire travel pathway by a vehicle from the point of origin to a destination (SAE International 2018).

Trust

The attitude that an agent will help achieve an individual's goals in a situation characterised by uncertainty and vulnerability (Lee and See 2004).

U

Uncertainty

This thesis recognises uncertainty as a system parameter that serves an indicator for capability and reliability.

Usability

Extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context (ISO 2018).

Usage specification

A particular level of driving automation within a particular ODD. Each feature satisfies a usage specification (SAE International 2018).

User

Person who interacts with the product (ISO 2018).

User experience

Person's perceptions and responses resulting from the use and/or anticipated use of a product, system, or service (ISO 2010).

V

Visual computing

The process of taking complex data and presenting it in a visual format that is easy to understand (Borojeni, Chuang et al. 2016).

Visual variable

Basic visual building blocks such as colour hue, position, size, or transparency (Bertin 1967).

APPENDIX A

Supplementary Content

A.1 Human-Automation Interaction

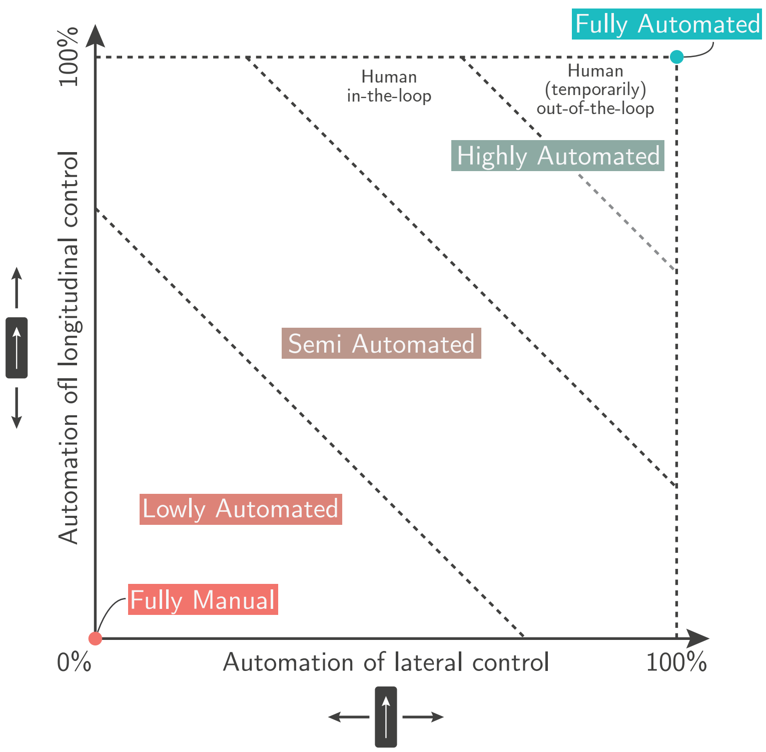


Figure A.1: Automation levels in relation to lateral as well as longitudinal guidance and control (based on Flemisch, Bengler et al. 2014, p. 344)

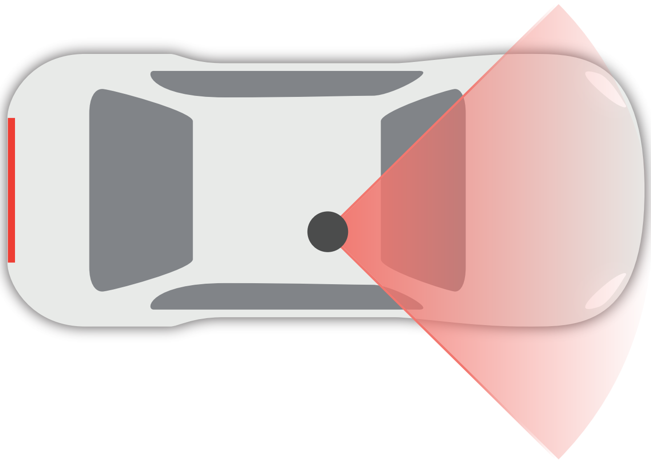


Figure A.2: Field relevant for driving (adapted from Kircher and Ahlström 2009)

Table A.1: Comparison between different models of SA (Stanton, Salmon, Walker, Salas et al. 2017, p. 12)

	Individual SA	**Team SA**	**System SA**
Defining features	Single person	Two or more individuals	Human and non-human agents
Typical features	Human constructed system	Human constructed system	Dynamic multi-agent systems view
Absent features	More than one person, non-human agents	System constraints and non-human agents	Internal information processing by individuals
Primary method	SAGAT probes	Information communication probes	EAST (Transactions)
Domain of origin	Aviation	Generic military	Maritime
Underpinning	Human information processing	Three-level model and team work theory	Perceptual cycle model, schema theory and distributed cognition
Definition	Perception of elements, comprehension of meaning and projection of future status	Shared understanding of a situation among team members at one point in time	Activated knowledge for a specific task within a system which relates to the state of the environments and the changes as the situation develops
Key citation	Endsley (1995b)	Salas et al. (1995)	Stanton, Stewart et al. (2006)
Citations	2253	482	239

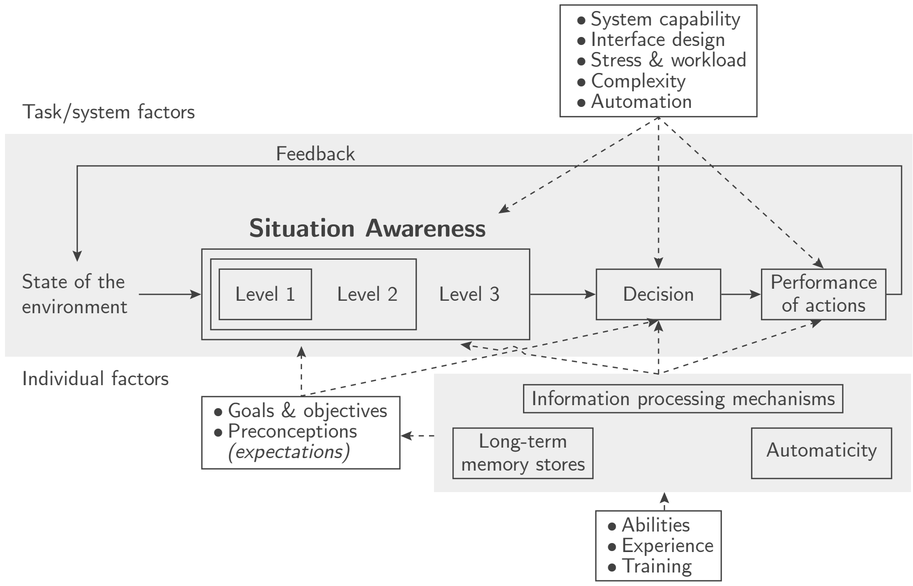


Figure A.3: Model of SA in dynamic decision-making (adapted from Endsley 1995b, p. 35)

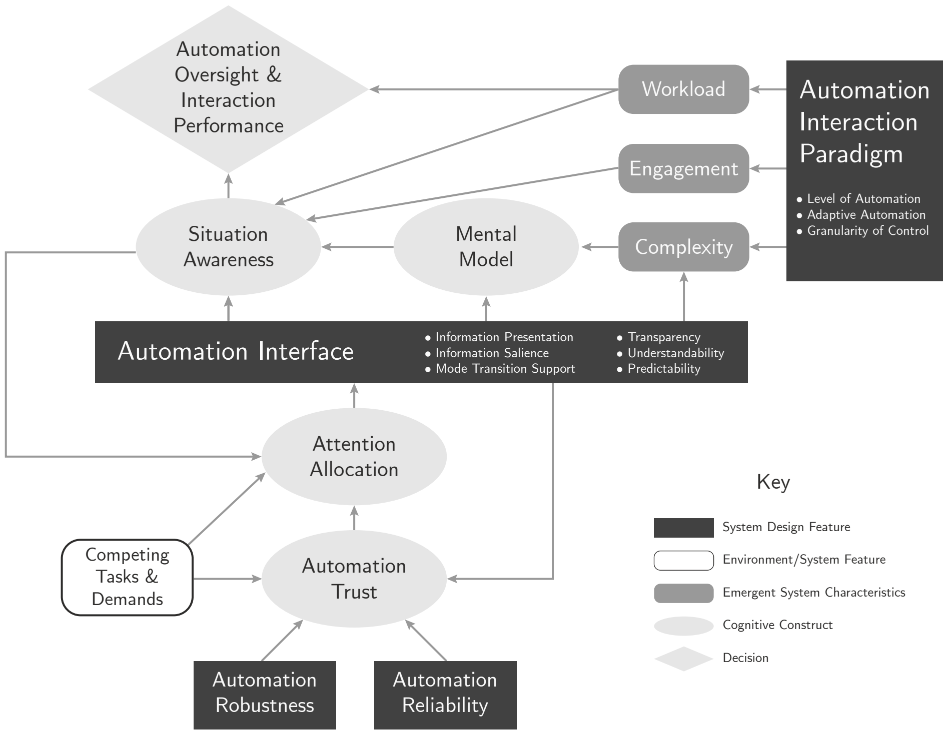


Figure A.4: HASO model depicting the relations of factors that influence oversight, intervention, and interaction with automated systems (adapted from Endsley 2017, p. 9)

A.2 Factors Affecting Vehicle Purchase Decisions

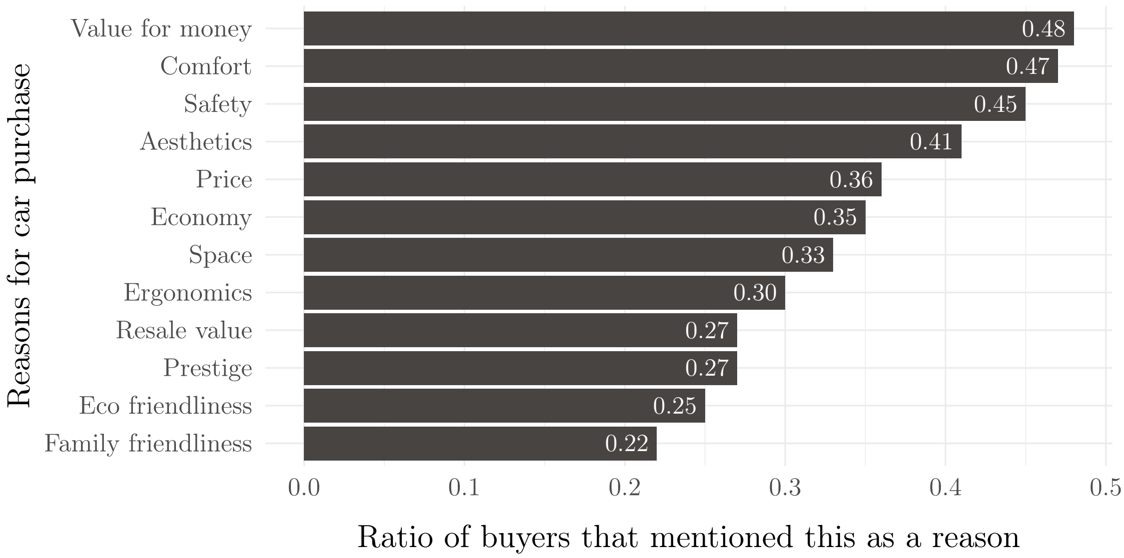


Figure A.5: Results from a survey involving 346 German car buyers that were asked what reasons determined their buying decision (Aral Aktiengesellschaft 2017)

A.3 Frameworks and Methods for Evaluating Peripheral Awareness Displays

This section provides supplementary information about the frameworks and methods for evaluating peripheral awareness displays presented in Chapter 5.

A.3.1 IRC Characterisation Framework

McCrickard, Chewar et al. (2003) propose the IRC characterisation framework to guide the development of notification systems. The IRC is based on the systematic establishment of critical parameters (Newman 1997) which represent user goals that were identified based on a survey of the literature. The following parameters were selected because of their measurability in user testing, accordance with cognitive processes, and manageability through design choices (McCrickard and Chewar 2003, p. 319):

- interruption: attentional shift evoked by the notification system
- reaction: speed and accuracy of user actions following display cues
- comprehension: recollecting and understanding the conveyed information later on

Combinations of *high* (1) or *low* (0) levels are possible for each parameter. For example, a peripheral awareness display for monitoring critical activities with the parameter values *high interruption*, *high reaction*, and *high comprehension* would result in IRC 111, which can be visualised using a cube (see Figure A.6). Each of the parameters is linked to the human information process (see Section 2.4). Thereby, interruption addresses working memory to perform the change of attentional focus. Comprehension leads to a flow of information through long-term memory to match the input with existing knowledge. Finally, reaction is the observable response of the user.

While this model provides a good starting point for characterising peripheral awareness displays, it lacks parameters to describe suchlike displays in detail. For instance, users may not be interrupted by a display but they are still aware of its existence. Further, a display may have varying degrees of interruption, depending on the urgency of the conveyed information. For instance, a visual baby monitor may become increasingly salient with the detection of more noise and movement from the nursery.

A.3.2 Toolkit for Managing User Attention in Peripheral Displays

Matthews, Dey et al. (2004, p. 248) propose a toolkit for peripheral awareness displays that aims to support three characteristics which the authors identified as crucial for managing user attention conditional on information importance: the abstraction of raw content, the

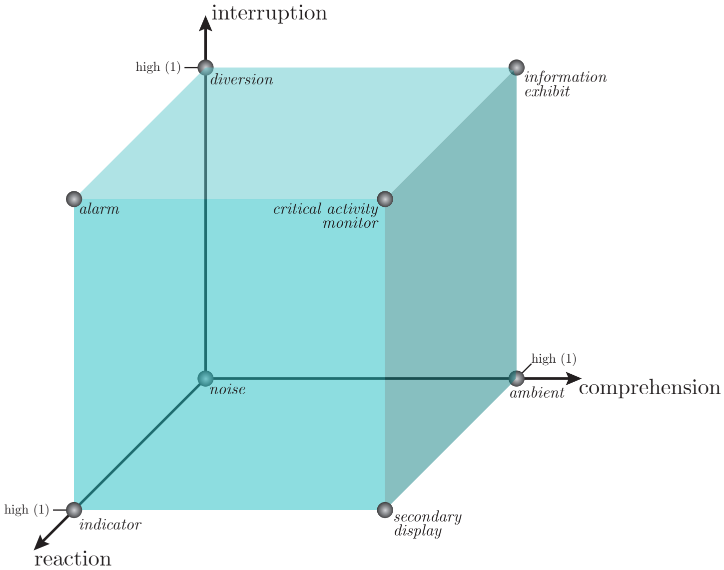


Figure A.6: Schematic representation of the IRC characterisation framework (adapted from McCrickard, Chewar et al. 2003, p. 321)

assignment of notification levels to content, and the design of transitions when content changes.

Abstraction aims to transform the raw data input to make content more accessible. To achieve this, key features of the input are extracted and/or the fidelity of the information is reduced. The colour hue of a light, for instance, may be mapped to the force that is exerted on passengers in an accelerating vehicle.

Notification levels delineate the relative importance of the communicated content. Higher notification levels indicate an increased urgency for the user to monitor the display. In contrast, lower notification levels correspond to less critical data and are displayed in a manner that does not actively attract user attention. As previously discussed, attention can be categorised in three different types (see Subsection 2.4.3): inattention, divided attention, and focused attention. Inattention refers to a state where objects cannot be consciously perceived. In divided attention, the attention is distributed among several perceptible objects. If all attentional resources are allocated to a signal object, this is referred to as focused attention. Based on this context, five distinct notification levels were defined by Matthews, Dey et al. (2004), ordered from low to high required attention:

1. ignore: information of no relevance that should not be displayed and does not correspond to any of the outlined types of attention
2. change blind (inattention): insignificant information that should not require attentional resources

3. make aware (divided attention): somewhat important information that requires few resources
4. interrupt (focused attention): significant information that requires the undivided attention of the user
5. demand attention (focused attention): highly significant information that requires the user to perform an action to stop the notification

Both *interrupt* and *demand attention* require the peripheral awareness display to briefly attract focused attention, thereby seemingly contradicting the terminology. However, as Matthews, Dey et al. (2004) emphasise, the capability to attract focused attention in critical situations allows the display to remain peripheral as long as the information is less critical.

Matthews, Dey et al. (2004) further discuss the transitions between different notification levels. As such, designers must evaluate how subtle or salient transitions must be to attract appropriate attentional resources. With increasing notification levels, the transitions are to be less subtle and more abrupt. For instance, low notification levels should transition with gradual and repetitive animations, such as fading. With increasing criticality, the displays should employ more and faster motion.

Concluding, the toolkit complements the IRC model by considering the attentional aspect of suchlike displays through the definition of notification levels and by adding the parameter abstraction. However, both presented models lack a consideration of the aesthetic emphasis of a peripheral awareness display.

A.3.3 Taxonomy for Ambient Information Displays

Pousman and Stasko (2006) pick up on the previously discussed framework and propose four design dimensions tailored to ambient information systems. The dimensions can be understood as design choices the developers and designers have to make. Analogous to the use of critical parameters presented in the IRC (see Subsection A.3.1), different displays can be categorised using levels ranging from low to high:

- information capacity: represents the number of discrete information sources that a system can represent
- notification level: the degree to which system alerts are meant to interrupt the user
- representational fidelity: describes a system's display components and how the data from the world is encoded into patterns, pictures, words, or sounds
- aesthetic emphasis: concerns the relative importance of the aesthetics of the display

The taxonomy adopts the notion of varying degrees in informational urgency (notification level) and abstraction (representational fidelity) that were previously presented by Matthews, Dey et al. (2004) (see Subsection A.3.2). However, it lacks the parameters relating to the comprehensibility of information and the users' reaction to display changes (see Subsection A.3.1).

Complementing the previously defined frameworks, the taxonomy considers the aesthetic emphasis of peripheral awareness display as well as the number of information sources that are used by the display. The former dimension can help designers to decide whether an aesthetically more pleasing design is to be considered a higher priority than a mainly functional design.

A.3.4 Activity Theory-Based Framework for Peripheral Displays

Matthews, Rattenbury and Carter (2007) propose a framework for peripheral awareness displays based on activity theory (Leontyev 1981). Activity theory can be defined as 'a conceptual framework based on the idea that activity is primary, that doing precedes thinking, that goals, images, cognitive models, intentions, and abstract notions like "definition" and "determinant" grow out of people doing things (Morf and Weber 2000, p. 81)'. In this context, individual actions are sought to be understood based on the analysis of activities. This leads to a hierarchical order whereby activities (top level) determine actions (intermediate level), and actions determine operations (low level). For example, assume an individual is building a house (= activity) (Kuutti 1996). This requires several actions, for instance to transport material to the house and subsequently assemble it to a roof. On the operational level, this requires the changing of gears when driving or the hammering of nails into wood. Thereby, *tools* are used to execute the actions, for example the truck she is driving or the gear stick with which she is changing the gears (Hashim and Jones 2007). In practice, 'activity theory is a valuable tool for researchers [...] as it enables a means of discovering human activity without the express explication of tasks by participants, instead, through the mediated study of the participant's tools an understanding of activity is revealed which includes tacit and explicit actions (Hashim and Jones 2007)'. Thus, activity theory aims to infer (otherwise hidden) human activities from their use of tools.

Grounded in the presented context, Matthews, Rattenbury and Carter (2007, p. 228) define activities as 'long-term projects of a person or group of people', whereby users perform activities in order to achieve their goals and satisfy their needs. The authors further argue that 'peripheral display designers need to be aware of their target user's activities and how a peripheral display could appropriately influence them'.

Whereas activity theory is of practical value for the identification and analysis of individuals and their (otherwise tacit) activities based on the examination of their tools (Hashim and Jones 2007), it is questionable if the evaluation of peripheral awareness displays warrants such a foundation. Specifically, Matthews, Rattenbury and Carter (2007, p. 231) point out that a peripheral awareness display supports 'at least one activity of its user and [...] is used primarily at the operation level rather than the action level'. The former aspect, support of an activity, is implied by the earlier dated definition of McCrickard, Chewar et al. (2003, p. 315), foremostly through the notion that the system communicates 'valued information', hence content that helps the user to some degree (= any activity). The latter part, operational level, is implied through the use of the term display itself. Considering the example above, monitoring any display can be considered an operation, except for the trivial situation in which monitoring the display has no underlying purpose. The principal benefit of peripheral awareness displays lies not in the fact that their use can be operationalised, but rather in that they use resources of the user that are not preoccupied (see Subsection 2.4.3). As such, the definition based on activity theory does not add any further content to the already existing definitions. The context of use as described in, for instance, ISO 9241-210:2010 (ISO 2010) (see Subsection 5.4.1.1) sufficiently considers the tasks operators are aiming to execute.

Matthews, Rattenbury and Carter (2007) continue with a specification of four different types of activities that a peripheral awareness display can support (Matthews, Rattenbury and Carter 2007, p. 230). As such, it is distinguished between *dormant*, *primary*, *secondary*, and *pending* activities. Dormant activities are trivial as they include any activities that are unlikely to be serviced by any tasks of the user in the specified setting. Primary activities are defined as those that are supported by the user's primary action. Secondary and pending activities are similar in that they are both monitored by the user without being at the focus of attention. Pending activities, however, are likely to be transformed into a primary activity soon, whereas secondary activities are to remain in the periphery of attention.

To describe how the different activity types affect peripheral awareness displays, Matthews, Rattenbury and Carter (2007, p. 237) provide several examples. For instance, an information ticker (= peripheral awareness display) helps an office worker with remaining aware of the arrival time of a bus she plans to take home while she continues to organise her belongings to get ready for the commute. According to the authors, the display supports the primary activity of managing her work-life balance.

This example highlights that the same display could be part of a secondary activity of another user. For instance, imagine an additional office worker is currently finishing up a presentation for a conference while monitoring the information ticker to remain aware of the bus he wants to take later. In this case, *advancing the career* could, for instance, be the primary activity, while the peripheral awareness display supports a secondary or pending activity. This highlights that it is more appropriate to describe peripheral awareness displays by the specific task (track bus arrival time) within the specified context (office) they are meant to fulfil rather than distinguish between different types of activities in the sense of activity theory, as these differ between individuals.

Further, Matthews, Rattenbury and Carter (2007) present three design dimensions for peripheral awareness displays: scope of use, class(es) of supported activities, and criticality. The latter refers to how important, or critical, the information is to the user. The former two design dimensions refer to the activities that a display supports. As these dimensions are grounded in activity theory, which was previously argued to be a controversial approach, scope of use and the classes of supported activities are not considered.

Apart from an activity theory-based framework, Matthews, Rattenbury and Carter (2007) present a set of five evaluation criteria for peripheral awareness displays that are based on interviews with designers of peripheral awareness displays, the authors' own experience, and previous research:

1. appeal (usefulness, aesthetics): refers to the qualitative enjoyment of a display; informed by Mankoff and Dey (2003)
2. learnability: refers to the temporal expense necessary for users to proceed to the effortless use of the display; informed by Matthews, Rattenbury and Carter (2007) and Shami, Lshed and Klein (2005)
3. awareness: refers to 'the amount of information shown by the display that people are able to register and use without focal attention (Matthews, Rattenbury and Carter 2007, p. 243)'; informed by McCrickard, Chewar et al. (2003)
4. effects of breakdowns: refers to 'how apparent breakdowns are to users and how easily users can recover from them (Matthews, Rattenbury and Carter 2007, p. 242)'; informed by Mankoff, Dey et al. (2003)
5. distraction: refers to 'the amount of attention the display attracts away from a user's primary action (Matthews, Rattenbury and Carter 2007, p. 244)'; informed by McCrickard, Chewar et al. (2003) and Shami, Lshed and Klein (2005)

Thus, Matthews, Rattenbury and Carter (2007) combine the previously presented frameworks and complement them by adding the criteria learnability and effects of breakdowns.

A.3.5 Heuristic Evaluation of Ambient Displays

The heuristics as selected and defined by Mankoff, Dey et al. (2003, p. 175), ordered by issue coverage and including definitions from Nielsen (1994b) are:

Sufficient information design The display should be designed to convey 'just enough' information. Too much information cramps the display, and too little makes the display less useful.

Consistent and intuitive mapping Ambient displays should add minimal cognitive load. Cognitive load may be higher when users must remember what states or changes in the display mean. The display should be intuitive.

Match between system and real world The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

Visibility of state An ambient display should make the states of the system noticeable. The transition from one state to another should be easily perceptible.

Aesthetic and pleasing design The display should be pleasing when it is placed in the intended setting.

Useful and relevant information The information should be useful and relevant to the users in the intended setting.

Visibility of system status The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

User control and freedom Users often choose system functions by mistake and will need a clearly marked 'emergency exit' to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

Easy transition to more in-depth information If the display offers multi-level information, the display should make it easy and quick for users to find out more detailed information.

'Peripherality' of display The display should be unobtrusive and remain so unless it requires the user's attention. User should be able to easily monitor the display.

Error prevention Even better than good error messages is a careful design which prevents a problem from occurring in the first place.

Flexibility and efficiency of use Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

In a case study about peripheral awareness displays for communicating information about e-mails, Matthews, Hsieh and Mankoff (2009) found that the ambient heuristics insufficiently consider the visibility of errors. In fact, Mankoff, Dey et al. (2003) noted this themselves as a shortcoming. Thus, such a heuristic should be added to accommodate error-related issues. For this, the following heuristic proposed by Nielsen (1994b) can be included:

Help users recognise, diagnose, and recover from errors Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

A.4 Guide to Using the PADE Framework

This step-by-step guide provides an overview how the PADE framework can be implemented to develop user-centred display designs

PADE framework
This icon represents which part of the PADE framework is being addressed by each step

PADE process
This icon represents which stage each step belongs to

Step 1:
Specification of the usage context
Peripheral awareness displays are typically not stand-alone devices but a component of a larger system. Knowledge of the usage context makes it easier for designers to consider the potential interplays of this system.

Who are the intended users of the display?
- Age, cognitive/physical abilities, ...
What user characteristics might limit the accessibility?
- Visual impairment, digital proficiency, ...
In which environment do users interact with the display?
- Definition of the wider system
What are the main tasks users perform while interacting with the display?
- Addressing the visual, auditory,... channel
What perceptual channels are least addressed by other tasks?

Step 2:
Specification of design dimensions
Categorising the display limits the scope of potential solutions and provides the basis for the weighting of evaluation criteria.

How important are aesthetics to intended users (rate from 1 to 5)?
- (1): non-relevant; (3): somewhat important; (5): main purpose
How critical is the information that is conveyed by the display?
- (1): artwork; (3): somewhat critical; (5): essential for survival
How many distinct pieces of information are to be conveyed by the display?
- (1): low, single piece of information; (5): high, >20
What is the degree of informational abstraction?
- (1): abstract; (3): iconic; (5): identical representations, images
To what degree can users personalise the display?
- (1): no personalisation; (5): configuration of information and communication method

---

Step 3:
Establish guidelines
Consider related work and derive recommendations that support the concept ideation process.

Step 4:
Concept ideation
Based on the previously established guidelines, develop display concepts under consideration of the usage context, design dimensions, and criteria

---

Step 5:
Heuristic evaluation
In a first step, evaluate the display in a low-cost heuristic evaluation to identify key problems with the created concept

(a) Describe your concept in an annotated visual presentation
(b) Provide evaluators with this presentation and a list of design heuristics (see Subsection A.3.5)
(c) Ask evaluators to identify problems with the display concept regarding each heuristic and to rate them in terms of severity (1-2: minor issues; 3: intermediate issues; 4-5: severe issues that discourage use)

Step 6:
Concept ideation
Improve the display concept based on the problems identified in the heuristic evaluation

---

Step 7:
Evaluation with the PADE questionnaire
Evaluation of the improved display concept using the PADE questionnaire within the intended context of use

(a) Define an appropriate scenario that replicates a realistic usage situation
(b) Recruit participants that fulfil the accessibility criteria
(c) Enact the scenario with participants
(d) Ask participants to fill in the PADE questionnaire (see Table 5.5)
(e) Calculate UX scores using Equation 5.3

A.5 Vibration Motors

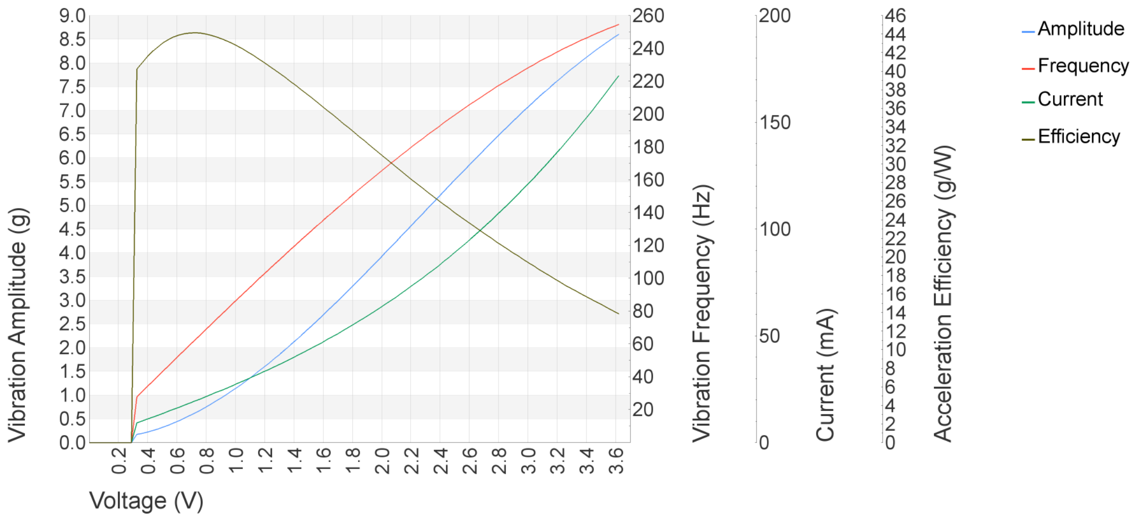


Figure A.7: Characteristics of eccentric rotating mass vibration motors used in experiments (Precision Microdrives Limited 2017)

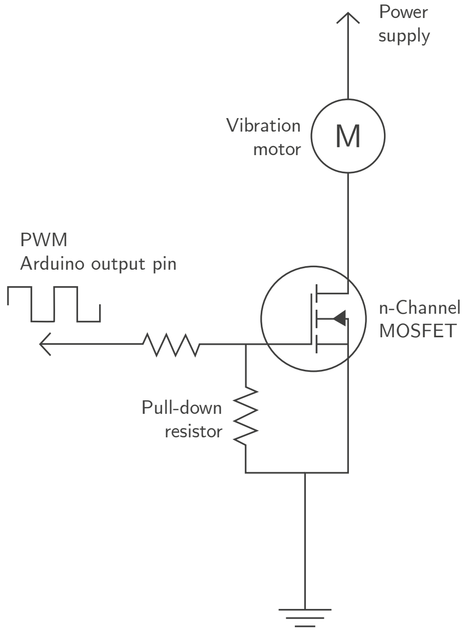


Figure A.8: Circuit diagram for control of vibration motors (Precision Microdrives Limited 2018)

A.6 Sample Interview Transcript

This section presents a randomly selected sample transcript in reference to the interview guide summarised in Table 8.1. The first three questions merely served introductory purposes and were not part of the analysis. Consequently, participant responses were only recorded for subsequent questions.

Experimenter: Tell me about how you experienced automated driving?

Participant: [No data recorded]

Experimenter: Tell me about the display in the instrument cluster in the last part of the study.

Participant: [No data recorded]

Experimenter: What differences did you notice compared with the display you used first?

Participant: [No data recorded]

Experimenter: Did your expectations regarding situations in which the system may fail vary between the two displays?

Participant: Yes.

Experimenter: How did your expectations vary?

Participant: Even though only one of the systems failed I instantly went to grab the steering wheel as well as did not trust the system, at all. With the singular one, even though I was not sure which part had failed it made me become conscious of all aspects rather than just one. I thought the steering would fail around courners which was something I was looking out for but did not expect speeding to fail. I thought the main issue however would be when cars pull out into the lane in front of you.

Experimenter: Having said that, would you rather use the system with one or two indications for uncertainty?

Participant: I prefer one indication for the overall uncertainty.

Experimenter: Tell me what you liked about this display.

Participant: Because it just refers to the system holistically, I would not trust another part of the system if one had failed. My instinct would be [...] to take full control rather than part control. It means I would also not have to think about the other one failing if I am controlling everything.

Experimenter: Do you have any further remarks?

Participant: No.

Following the procedure outlined in Table 3.2, the individual meaning units were increasingly abstracted (see Table A.2). As the participant indicated that the expectations differed and that the system-wide uncertainty display was preferred, the codes were added to the bottom left box of Figure 8.6.

Table A.2: Abstraction of interview responses

Condensation	Code	Category
looking out for failures in certain situations	Guides focus	Attention allocation
would not trust rest of system if one part fails	Trust contagion	System-wide trust
instinct is to take over the complete driving task	Completely takeover	Decision selection and action implementation
no additional thinking about other failures required	Reduce workload	Workload

APPENDIX B

Supplementary Data

B.1 Supplementary Data for Chapter 4

Table B.1: Descriptive statistics of subjective trust indications depending on visibility range

Group	**Visibility range** [ft]	**Subjective trust** M [%]	SD [%]
Control	1300	78.706	14.110
	2300	72.309	18.751
	3300	76.804	15.436
Experimental	1300	42.941	25.190
	2300	63.206	19.400
	3300	76.667	14.549

Table B.2: Descriptive statistics of the NDRT solving rate depending on visibility range

Group	**Visibility range** [ft]	**NDRT solving rate** M [min^{-1}]	SD [min^{-1}]
Control	300	27.497	9.239
	1300	28.376	6.971
	2300	27.787	7.530
	3300	28.707	7.760
Experimental	300	14.364	7.874
	1300	17.153	7.058
	2300	19.628	7.337
	3300	24.610	6.945

Table B.3: Descriptive statistics of the monitoring frequency regarding each section

Group	**Section**	**Monitoring frequency** M [s^{-1}]	SD [s^{-1}]
Control	TOR-80-40	0.106	0.133
	TOR-40-0	0.101	0.121
Experimental	TOR-80-40	0.275	0.185
	TOR-40-0	0.487	0.201

Table B.4: Descriptive statistics of the monitoring ratio regarding each AOI and uncertainty level for both display types

Section	**AOI**	**Group**	**Monitoring ratio** M	SD
TOR-80-40	FRD	Control	0.102	0.116
		Experimental	0.138	0.082
	Instruments	Control	0.006	0.010
		Experimental	0.004	0.013
	NDRT	Control	0.454	0.315
		Experimental	0.272	0.246
	UD	Control	0.000	0.000
		Experimental	0.045	0.030
TOR-40-0	FRD	Control	0.112	0.129
		Experimental	0.344	0.196
	Instruments	Control	0.008	0.017
		Experimental	0.012	0.016
	NDRT	Control	0.452	0.295
		Experimental	0.065	0.094
	UD	Control	0.000	0.000
		Experimental	0.114	0.086

B.2 Supplementary Data for Chapter 7

Table B.5: Descriptive statistics of the monitoring ratio regarding each AOI and uncertainty level for both display types

Uncertainty	AOI	Display	Monitoring ratio *M*	*SD*
Very low	FRD	Peripheral	0.045	0.053
		Digital	0.055	0.049
	NDRT	Peripheral	0.594	0.322
		Digital	0.552	0.326
	UD	Peripheral	0.003	0.008
		Digital	0.020	0.020
Low to moderate	FRD	Peripheral	0.164	0.157
		Digital	0.122	0.097
	NDRT	Peripheral	0.456	0.321
		Digital	0.349	0.218
	UD	Peripheral	0.041	0.032
		Digital	0.076	0.070
Moderate to high	FRD	Peripheral	0.359	0.242
		Digital	0.264	0.165
	NDRT	Peripheral	0.211	0.257
		Digital	0.227	0.218
	UD	Peripheral	0.026	0.028
		Digital	0.082	0.070
Very high	FRD	Peripheral	0.451	0.165
		Digital	0.318	0.119
	NDRT	Peripheral	0.043	0.068
		Digital	0.081	0.114
	UD	Peripheral	0.011	0.005
		Digital	0.076	0.046

Table B.6: Descriptive statistics of the NDRT solving rate regarding each uncertainty level for both display types

Uncertainty	**Display**	**Subjective trust** M [%]	SD [%]
Very low	Peripheral	87.698	15.475
	Digital	85.646	15.780
Low to moderate	Peripheral	73.646	16.042
	Digital	71.396	15.876
Moderate to high	Peripheral	48.333	21.803
	Digital	50.542	22.716
Very high	Peripheral	29.000	26.721
	Digital	33.208	25.505

Table B.7: Descriptive statistics of the NDRT solving rate regarding each uncertainty level for both display types

Uncertainty	**Display**	**NDRT solving rate** M [min^{-1}]	SD [min^{-1}]
Very low	Peripheral	30.855	5.205
	Digital	29.868	6.297
Low to moderate	Peripheral	25.221	9.065
	Digital	26.176	7.426
Moderate to high	Peripheral	21.104	9.930
	Digital	20.880	7.842
Very high	Peripheral	8.598	8.447
	Digital	10.849	10.159