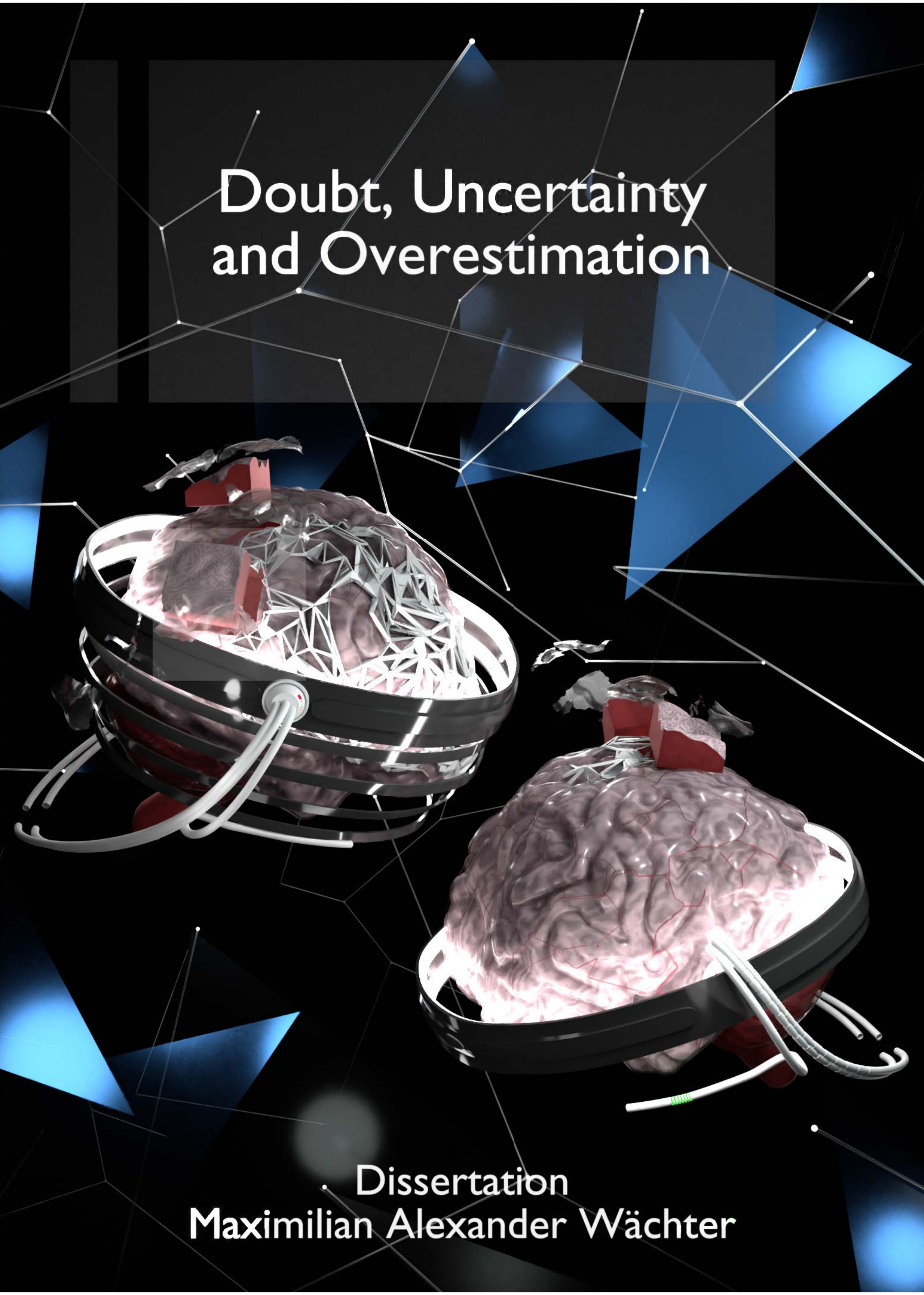
Doubt, Uncertainty and Overestimation



Dissertation

Maximilian Alexander Wächter

Doubt, Uncertainty and Overestimation

relations between humans and machine agents in virtual reality experiments

Maximilian Alexander Wächter

Supervised by Prof. Dr. Peter König and Prof. Dr. Gordon Pipa

Neurobiopsychology and Neuroinformatics

Institut of Cognitive Science

University of Osnabrück

December 2021

Dissertation zur Erlangung des
Grades Doktor der Naturwissenschaften (Dr. rer. nat.)
Tag der Disputation 01.02.2023

Supervisors:

1st Supervisor: **Professor Doctor Peter König**
University of Osnabrück

2nd Supervisor: **Professor Doctor Gordon Pipa**
University of Osnabrück

3rd Supervisor: **Doctor Sabine König**
University of Osnabrück

RTG Mentor: **Gabriela Pipa**
University of Osnabrück

Acknowledgements

First, I would like to express my deepest gratitude to Prof. Dr. Peter König and Prof. Dr. Gordon Pipa for their outstanding supervision and advice during the time of the Ph.D.: for the invested time, the motivating support, and a sheer unlimited willingness to enable the projects that will be part of this Thesis. I am more than thankful for the opportunity to learn and grow in the environment that I was part of during the last three years. I could not have wished for a better environment.

This also accounts for the additional supervisors and mentors, Dr. Sabine König and Gabriela Pipa. Thank you for your countless efforts to support us as a working group and for the personal consultancy. Thank you for your critical thoughts and your corrections all the times we went astray.

In line with this, I would like to thank Farbod Nosrat Nezami for the last adventurous four years of our joint project, Westdrive. Thank you for your support, your fabulous ideas, your endless enthusiasm, and your friendship. Without you, none of the projects would have come to existence.

I also would like to express my gratitude to the Stiftung Stahlwerk Georgsmarienhütte for granted scholarship. The funding I received enabled me to focus on the research projects and was thus an essential prerequisite for the success of the projects presented here. Here, I would also like to thank Fanatec and the Endor AG for the generous hardware sponsorship

Special thanks to all the people of Team Westdrive who accompanied me during the last years and the fantastic time we spent together on the different projects. For all the ideas, persistence, and commitment I would like to thank:

Johannes Maximilian Pingel, Philipp Spaniol, Anke Haas, Lea Maria Kühne, Lynn Keller, Sumin Kim, Frederik Nienhaus, Thomas Klein, Nora Maleki, Deniz Gün, Linus Tiemann, Stefan Balle, Lucas Essmann, Richard Ruppel, Hristofor Lukanov, Marc Vidal dePalol, Shadi Derakhshan, Tracy Sánchez Pacheco, Ashima Keshava, Vincent Schmidt and Dr. Sylwia Kaduz - and many more people from the Neurobiopsychology and the Neuroinformatics of the IKW.

I also would like to thank my friends Jan Dirk Capelle, Jonas Taubmann, Etienne Herrmann and Mattis Osterheider for their open ears and minds during the last three years.

Especially, I would like to thank my Fiancée Nicola for her never-ending patience and support in everything I did.

Abstract

Artificial intelligence is becoming ubiquitous in our everyday lives with the emergence of powerful machine learning methods and the ever-increasing amount of available data. One can find "intelligent" machines in operating theaters, factories, in almost every pocket, and even space. The next milestone will be the fully autonomous vehicle. With this technology, artificial intelligence is not only an abstract term or bound to cyberspace; it is moving significantly closer to humans, functioning as a collaboration partner. Besides the remaining technical issues, autonomous robots raise questions on social, ethical, and legal issues that we have to evaluate to unleash the full potential of autonomous systems as artificial agents.

These studies include new challenges of human-machine interaction, trust and acceptance of self-driving cars, and ethical issues associated with programmed non-human behavior. In addition to the experiments themselves, this thesis also includes the VR toolkit created to develop the experiments. The toolkit itself provides a foundation for further research in virtual reality. By way of introduction, we will first clarify the extent to which AI impacts vehicles. Here, an insight into the definition of the different automation levels of modern vehicles is given. This thesis then addresses various research questions of the respective automation level in ascending chapter order. The start will be the investigation of take-over requests in the context of highly automated but human-supervised driving. Next, the virtual reality studies investigate how human behavior in highly automated driving can be optimized using a human-centered design. Since data analysis is ongoing, this section will give preliminary results but focus on the resulting toolkit as a virtual environment with custom functionalities. In chapter three, we will shift focus to higher automated cars, where the driver cannot take over control instantly. Here we developed a self-explanatory virtual agent to increase trust and acceptance. In chapter, four we tested the different communication strategies in a large-scale virtual reality experiment. We found that demographic factors influence acceptance more than what the vehicle does or communicates. In the final study of this thesis, we examine human attitudes toward fully autonomous vehicles. Here we examine how people behave in a potentially dangerous situation in a self-driving vehicle and whether users in a moral dilemma decide according to deontological ethics or whether their decisions can be grouped under utilitarian ethics. We then use the results of this study to develop normative ethics for self-driving vehicles.

The results of the studies will be summarized to create a unified concept of a human-centered interaction design. It is, though, to increase trust and acceptance and ultimately, through clever algorithms, allow for human performance

to be increased in such a way that legal and ethical problems can be solved. This will enable the numerous promises of autonomous mobility, such as integrating people with impairments or significantly reducing emissions in traffic.

Contents

List of Figures

List of Tables

List of Abbreviations

NPCs non-player characters 66
TAM technology acceptance model 79
AVAS anthropomorphic voice assistant system 82
MANOVA multivariate analysis of variance 84
ANOVA analysis of variance 84
LDA linear discriminant analysis 85
HSD Tukey honest significant difference 85
DALY disability-adjusted life years 127

1
From Data to Driving

1.1 | Times of Automation and Artificial Intelligence

Within the last decade, artificial intelligence (AI) has become part of our everyday lives. Simply put, artificial intelligence is the ability of machines to learn from data by analyzing it to make predictions for the real world (Ongsulee, 2017; Warwick, 2012). The requirements for this are surprisingly simple: It only needs a large amount of data and a pattern in the data that cannot be solved purely mathematically. Due to these simple requirements, the possible areas of application are almost unlimited. Powerful machine learning (ML) methods for analyzing all kinds of data and potentially learning from it surround us: They are applied in medicine, social media, banking and investments, customer services, robotics, logistics, and science. Since a large part of society and industry are in an ongoing process of digitization, the connection between humans, devices, and data will increase further. Algorithms will likely be able to take over the task of driving. However, this task is complex and deeply human, and therefore difficult for the machine to learn. The foundations and open challenges of this field of research are the topic of this thesis since legal and ethical issues stem from how machines learn and perceive. Exactly that is the content of this introduction.

Machine learning is inherently fundamentally different from that of classical computer programs as a human-defined set of rules. Instead, the machine extracts features of the data by itself, guided only by simple instructions. Such a procedure comes in handy in complex tasks which cannot be pinned down mathematically, such as assigning a human face to a matching age and gender. With classical programming, a large and potentially impossible set of rules would have to be defined. Instead, artificial systems usually use an extensive data set of already classified data and learn by themselves over many iterations. Neural networks are often chosen for this, a subset of machine learning called deep learning. The "depth" in this term stems from the networks structure of node layers, weights, and thresholds. Just like in a human brain, input is calculated, and if a specified threshold value is reached, the node in the network is activated, passing information to the next layer of the network (Figure 1.1). So the knowledge of that system lies within the structure of the network. However, this means that the computer's information is usually inaccessible and not explainable to humans. What happens between the input and the output of such a system occurs within a "black box." This is due to the logic being inferred automatically from large data sets, not explicit rules.

Thus, if the input data is not checked correctly, the system can learn human errors or biases. Depending on the usage and scale of a possibly faulty network, this could have far-reaching consequences. One example is the increasing use of predictive policing, where poor neighborhoods are specifically targeted because of a problematic connection of spatial location and criminality (Shapiro, 2017). In order to derive the most benefit for humankind from artificial intelligence, it is crucial to create transparent or at least explainable systems whose decisions are discernible and accepted by humans.

However, before this problem is addressed, the term artificial intelligence should be defined. John McCarthy, one of the field founders, defines AI as "[...]the science and engineering of making intelligent machines, brilliant computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to biologically observable methods" (McCarthy, 2007). Some definitions even include a system to act or decide in a humanlike manner (Ghahramani, 2015; Hirschberg & Manning, 2015; Rudin & Wagstaff, 2014). However, the definition of the term machine intelligence is not simple, as machines are still a fair way from encompassing humanlike behavior, reasoning, and decision-making in their entirety (Monett et al., 2020). In most cases, machine learning refers to particular areas or a single specific task, as general artificial intelligence still represents a challenge, and it is even doubted that a "strong" and thus humanlike artificial intelligence will ever exist (Martinez, 2019).

Nevertheless, AI in its narrower and more specific versions affects our society dramatically (Rudin & Wagstaff, 2014; Verma, 2019). Over the last decades, marketing changed drastically, as customers conveniently receive personalized advertisements according to personal needs and interests based on previous choices (Addagarla & Amalanathan, 2020; Cohen et al., 2020). Artificial intelligence fundamentally changed work with databases, optimizing and automatizing all processes involved, since it is used to act faster to cope with masses of data that a human would not be able to even screen in more than a lifetime (Bose & Mahapatra, 2001; Paschek et al., 2017). It is also able to revolutionize health care with faster, more reliable diagnostics and improved prophylaxis and treatment evaluation (Check Hayden, 2014; Davenport & Kalakota, 2019; Yu et al., 2018). However, behind these spectacular achievements of artificial intelligence, a darker side should be addressed.

Put succinctly, the misapplication of mathematical models and algorithms can pose a threat to democracy and freedom (Verma, 2019; Wirtz et al., 2020). Due to the mass of available data and the potential lack of data protection (Hwang, 2015), people's behavior can be predicted quite accurately and manipulated accordingly, e.g., by allowing only certain information to reach them or by providing extra filtered information (Bovet & Makse, 2019; Cybenko & Cybenko, 2018). Likewise, people are misclassified and often discriminated against on the basis of faulty or opaque models with abstract parameters (Sloane & Moss, 2019; Tian et al., 2021). Error, and therefore discrimination, does not necessarily have to lie within data or the algorithm itself but can be introduced into predictions from outside. Sometimes by mistake or an adversarial attack, where minor changes in the input data change the output dramatically (Carlini & Wagner, 2017; F. Chen et al., 2020; Waldrop, 2019).

AI techniques are a vector of innovation, but they are also a challenge in cybersecurity, safety and explainability. In current applications, artificial intelligence is often based on multilevel, multilayer, entwined algorithms and models, like artificial neural networks. Consequently, decisions made by the machine cannot be inspected and controlled, since the human does not know how the machine arrives at a certain judgment (Carlini & Wagner, 2017; Georgevici & Terblanche, 2019). The lack of explainability may not be a problem in an application, e.g., automated reordering for inventory or analyzing a football game. However, the more power an algorithm has over humans, like in the case of criminal justice, the more disruptive the lack of transparency and explainability becomes [1] (Ebers, 2021; EU commission, 2020; T. Miller, 2019).

Here, it can be objected that biological neurons in the human brain are also not transparent in decision-making. While this is true, humans, especially with profes-

[1]Explainability and transparency are often used as synonyms. However, the distinction between them is important. While explainability only means explaining post hoc how a decision is made, transparency implies a deep understanding of the system that can reveal every step, e.g., an assessment. Accordingly, transparent systems are more challenging than explainable systems.

sional training, learn from explicit rules rather than only provided data. To shed more light on the problem of making automated decisions more comprehensible, three subproblems are pointed out. The first concerns the implementation. This means that it is clear which technical principles are used, i.e., which operating model is chosen with which coefficients, weights, and thresholds. These models are often used in science and are called "white-box" models (Loyola-Gonzalez, 2019; Rudin, 2019). The second subproblem concerns the specification, i.e., the communication of which methods are applied in which context and especially which training data is used. This form of traceability enables replication and is also prevalent in science. In industrial applications, however, this knowledge is often hidden as it is the intellectual property of the company or too complex to be computed afterward. An example of this is Instagram's algorithm (O'Meara, 2019), the Facebook news feed (Horwitz, 2021), or the Spotify recommendation algorithm (Werner, 2020), each being the crucial success factor for the company. Finally, an example of a highly complex network is the OpenAI GPT-3 network with over 150 billion parameters (Dale, 2021).

The third subproblem is transparency and is linked to the understanding of underlying mechanisms. Transparency is defined by how well a human can understand the decisions of a system (Doshi-Velez & Kim, 2017). The idea is to extract elements of the network that allow the human to understand the outcome without using a formal definition (T. Miller, 2019). It may also includes a possible demonstration of how the the algorithm follows certain specifications. Up to now, there is no AI system that is able to achieve full transparency (Hamon et al., 2020).

In general, it is possible to disclose all parts of the algorithms. The disclosure includes the features and the data the system used for training. The disclosure also includes the model and thresholds, if any. However, since, as mentioned before, the algorithms that influence our lives are not accessible because they are intellectual property or are high dimensional models, any attempt to do so is reverse engineering of these black-box models (Lipton, 2018). Another way is to make particularly prominent features visible, that is, to highlight the data that have the most substantial impact on the outcome. Similar to this is the possibility to give counterfactual explanations to identify features that have a decisive impact on the outcome (Adadi & Berrada, 2018). Since these methods are based on statistics, they are susceptible to uncertainty and error. The result of both variants of reverse engineering, be it global or one of the features, is nevertheless mostly an incomplete surrogate model, which does not represent the performance of the model it is replicating (Guidotti et al., 2018; Lipton, 2018).

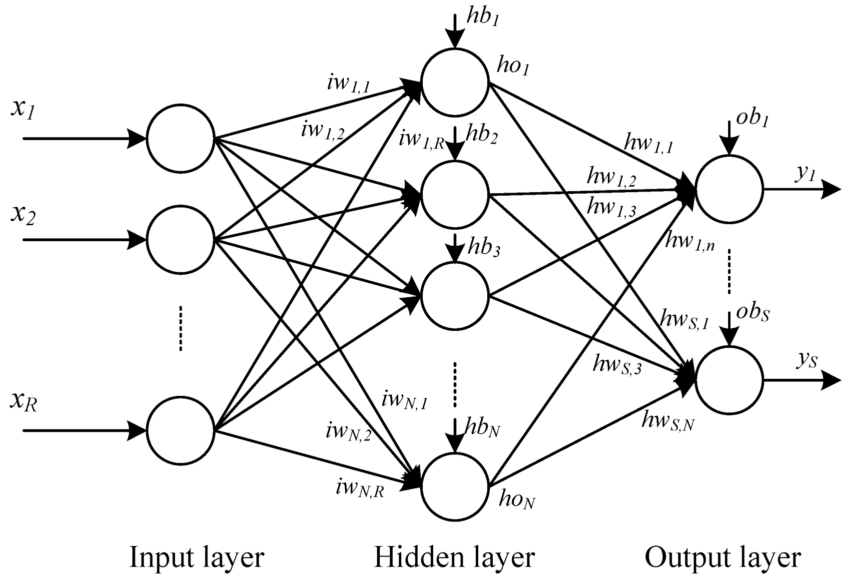


Figure 1.1: An example of a artificial neural network with a single hidden layer, referred to as a black box. These networks can grow to multiple hidden layers and millions of weights (J.-F. Chen et al., 2014)

The scientific community is aware of these dangers, and there is a call for a more white-box models. These models are designed to provide reliable and easy-to-understand explanations for prediction (Rudin, 2019). The problem, however, is that it is not always clear whether it is possible to develop a designated interpretable model with the desired accuracy. Even within the scientific discourse, it is still controversial to what extent these transparent models can be applied since often, then the performance of black-box models, e.g., deep neural networks, is much better than that of white-box models. However, this problem and the closely related issue of reliability are part of an ongoing scientific debate as to whether there is a trade-off between explainability and accuracy, depending on the context of use or limitations. However, this discourse will not be referred to further because the use case of highly automated and fully autonomous driving is, in any case, a highly complex and human task that, however, in principle permits explainability (J. Kim & Canny, 2017).

Just as the scientific community, the European legislature has been addressing the question of how to take advantage of this technology without incurring the far-reaching consequences of faulty models (Cohen et al., 2020; Dudley & Wegrich, 2016; EU commission, 2020). This becomes particularly clear when looking at the goals of transparent AI of the high-level expert group of the European Commission EU commission, 2020. The groups states that the pillars that should lead to trust and acceptance are priority of human action, technical safety, privacy, transparency in explainability, fairness, social and environmental well-being, and accountability (EU commission, 2020). However, it is unclear which standards should be used to evaluate a system concerning its transparency and comprehensibility (Kozuka, 2019; Veale, 2020). On the one hand, this means a degree of uncertainty about potential use cases, e.g., in medical applications. In these scenarios, treatments and diagnostic methods require legal validation (Kozuka, 2019; Shaw et al., 2019); on the other hand, it means significant freedom in research as a broad framework is available for researching measures in the absence of legal requirements. These problems of artificial intelligence in general, however, will not be explored further here. Instead, they are only a first indication of how deep the complexity of acting machines extends. Later in this thesis, we will come back to this subject regarding trust and acceptance of humans in autonomous systems.

The foundation needed to achieve self-driving systems is artificial intelligence and machine learning (Grzywaczewski, 2017; Hussain & Zeadally, 2019). How vehicles perceive their environment and react to it is so complex that the vehicle chassis and mechanics almost recedes into the background. Thus self-driving vehicles can be practically described as computers with wheels — where software development is as important as, if not more so than, for traditional business units (Alt et al., 2020). This is due to the fact that the code programmed into the vehicle and the vehicle's associated capabilities already significantly influence purchasing decisions in many cases - and this phenomenon is likely to intensify in the future (Dajsuren & van den Brand, 2019).

In addition, to sensing traffic and controlling behavior, the self-driving vehicle is an interaction partner with an occupant (Maurer, 2015; Puertas-Ramirez et al., 2021). This interaction opens up another entirely different line of research, the human-machine interaction (HMI). This independent field of study has its own set of open questions in the context of individual automated traffic, where a framework for active human-machine interaction in AVs is needed (Bengler et al., 2020). Here, the relation of transparency in decision-making to self-driving vehicles becomes clear: This is a significant factor for not only increased trust and acceptance among potential users, but also legal determinations in the event of an accident (Cysneiros et al., 2018; Gillmore & Tenhundfeld, 2020).

To summarize, this thesis provides an overview of the potentials and dangers of self-driving vehicles with different levels of automation. In particular, it covers the problems arising from human-machine interaction. Therefore, this work will present an experiment on human-machine interaction is examined in partially automated cars with control takeover. In multi-modal conditions, it is investigated whether a more controlled and safer takeover can be achieved with different warnings. Following this, we will see how requirements change with the introduction of fully automated vehicles. Here a way to increase trust and acceptance via different measures of communication between the car and the human driver will be proposed. Regardless of what a decision of a self-driving vehicle might be, the decision should be at least partially explainable regarding how the machine reached that conclusion or chose that action. That is the content of the third and fourth chapters of this thesis. While chapter 3 details a virtual reality toolbox created as an open-science project to enable human-machine interaction in the context of self-driving cars, chapter 4 describes the study conducted using this virtual toolkit and an additional survey after the experiment. Here, this thesis will show that different communication strategies in a fully autonomous vehicle make a difference in trust and usefulness. Nevertheless, we found that demographic factors, such as gender and age, affect the attitude toward self-driving cars more significantly than in-car communication.

Subsequently this thesis will deal with the decision-making of self-driving vehicles. After all, should fatality or severe injury occur due to error, regardless of the cause of the accident, the car will have to make a decision. What this might look like and the problems with non-human programming decisions will be discussed in chapter 5.

Before we look closely at the human-machine interaction in self-driving vehicles and the associated issues, we will examine the technology of self-driving vehicles in more detail in the following section. Here we will glance at the functionality and the not so trivial definition of automation levels and the open legal questions of autonomous vehicles. After this, the focus will shift towards the social implications of self-driving cars on our society. Additionally, we will discuss why acceptance of, and trust in, this technology are of utmost importance for its realization. Finally, the challenges this poses for human-machine interaction will be addressed.

1.2 | Mechanical Machines as Autonomous Agents

For decades, automobile engineers have been trying to create a fully self-driving vehicle (Takacs et al., 2018). Since the invention of the car in 1886, there has been a constant progression of technology.[2] However, the car has never really managed to live up to its own definition. The term "automobile" is a combination of the Greek word "autós" for "self" and the Latin "mobilis" for "moving" (Dietsche & Kuhlgatz, 2014). In the true sense, this definition only applies to the engine itself once it has been started because a car alone would not drive or steer without a human operating it.

The idea of the self-driving car originated in the United States in the early 1920s (Nguyen, 2019). In this decade, 365,000 US citizens were affected in motorized traffic accidents annually (Norton, 2008). Just as today, human driver error was identified as the primary cause of these injuries and deaths (Norton, 2008; Stanton & Salmon, 2009). In response, ideas of self-driving cars followed, ideas that, whether for reasons of cost or lack of technical solutions, have been discarded. In the days of mass motorization in the 1950s, there were many creative ideas and visions of what automated transport could look like in the future, and there was heavy investment in technology for autonomous driving vehicles (ADVs) (Kröger, 2015). Due to the oil crisis in the 1970s, automakers put aside plans for ADVs since costly investment seemed too uncertain. Thus, revolutionary ideas became evolutionary ones and incremental development set in as assistance systems were introduced to the market (Campbell et al., 2010; Kröger, 2015). The first technical driver-assist innovation in the mid-1970s was the ABS (anti-lock braking system) (Demel & Hemming, 1989). This system marked a first milestone for the automation of modern cars as it improved human driving behavior through controlled braking mechanisms to an extent that human drivers are not capable of. Since then, more and more functionalities in vehicles have been automated: parallel parking (Pohl et al., 2006), automated lane keeping for highways (Shinq-Jen Wu et al., 2005), and automatic distance control (Swaroop & Hedrick, 1999).

Now, after more than 100 years of development, the fully autonomous car seems, at least technically, within reach: a vehicle that can truly drive without a human actor or even attendant. This development appears to be reasonable as the benefits of fully automated traffic are thought to have an immense impact on our society and

[2] Excitingly, the automobile has undergone many changes in its history from its design to its safety features – and with the help of modern computer systems – communication and diagnosis systems. Still, the fundamental drive technology of internal combustion has remained the same for over 100 years (Dietsche & Kuhlgatz, 2014)

the environment. These impacts include a drastic decrease in accidents and traffic-related fatalities (Bergmann et al., 2018; Hevelke & Nida-Rümelin, 2015a, 2015b), a significant reduction in emissions of CO_2 and other greenhouse gases (Chehri & Mouftah, 2019), less congestion, and inclusion of people with limited access to personal transport, such as the elderly or people with disabilities (Eby et al., 2016; J. Yang & Coughlin, 2014).

One cannot deny the advantages of this technology. However, in addition to the technical challenges, there are a large number of legal and ethical issues that need to be addressed (Faulhaber et al., 2019; Kallioinen et al., 2019; Sütfeld et al., 2017). For this reason, it is not realistic to say precisely when autonomous vehicles will be able to drive completely autonomously on our roads (Gessner, 2020; Shalev-Shwartz et al., 2018). Even beyond this, the shift from a driver-centered culture, where the driver is in complete control at all times, to a "being-driven" culture will undoubtedly be difficult (Bjørner, 2019). The reason is that car automation is not comparable to other automated objects in the past: Unlike a washing machine, a self-driving car does not take on unpleasant, exhausting work, but rather takes on an exciting and often joyful, albeit risky, activity (Hagman, 2010; Wedlin et al., 1992).

However, the fully autonomous car is possible only if a human driver is no longer necessary to monitor the driving operation, nor can the human intervene in the driving operation at all. Such a vehicle must find its way around in traffic on its own and communicate with other road users, both human and machine. Furthermore, it must be able to weigh risks and make legally protected decisions even under sub-optimal conditions, which are then legally protected (Shadrin & Ivanova, 2019). Only when this is guaranteed can a car be said to be a fully autonomous agent. Furthermore, the self-driving vehicle fulfills the definition of the automobile and the utopia of environmentally friendly and low-risk individual transportation.

Indeed, the problem of defining the driving task opens up one of the most challenging problems in research on human cognition: how to model such complex systems as human behavior behind the steering wheel (Karwowski, 2006). Without going much further into the material, this should provide an insight into how complex the intuitively simple task of driving a car is. Driving as task consists of a variety of traffic related actions and decisions, which can be divided into three different sub-groups of tasks: strategic efforts, tactical efforts and operational efforts (Michon, 1985) (Figure 1.2, right side). Strategic tasks represents the highest level in this model. This task generally include planning the journey ahead. This includes planning the route, i.e. when and why exactly where to travel. These plannings are made in the long term, since they concern the entirety of the route.

The tactical efforts collect decisions to perform maneuvers in traffic that involve other road users, such as when and how to overtake another vehicle, when to change lanes, or even how to adjust the speed to the current traffic situation. The tactical efforts have to meet the objectives of the strategic level. Since these efforts involve reactions to specific situations, they happen within seconds.

The last are the operational actions, which are very short term (in milliseconds) and sometimes even performed unconsciously. These actions represent the execution of the higher level objectives, for example micro corrections at the steering wheel to hold course, acceleration or braking during a simple controlled drive to keep the road position, as well as all emergency interventions to compensate for sudden events.

Rasmussen (1986) proposed a compatible model of driving activity at three levels (Weller & Schlag, 2007). This is not concerned with the driving task as such, but with the cognitive mechanisms involved. As depicted in Figure 1.2 this model distinguishes between knowledge-based activities, rule-based activities, and skill-based activities. Knowledge-based activities, are those that involve deliberate reflection and promote a transfer of knowledge. Thus, it is possible to solve problems that have not been encountered before. Rule-based activities are actions where a decision rule already exists and corresponds to a defined situation. Here, a rule is selected and acted upon. Skill-based activities do not require active deliberation. In their case, a reaction to a known procedure follows automatically.

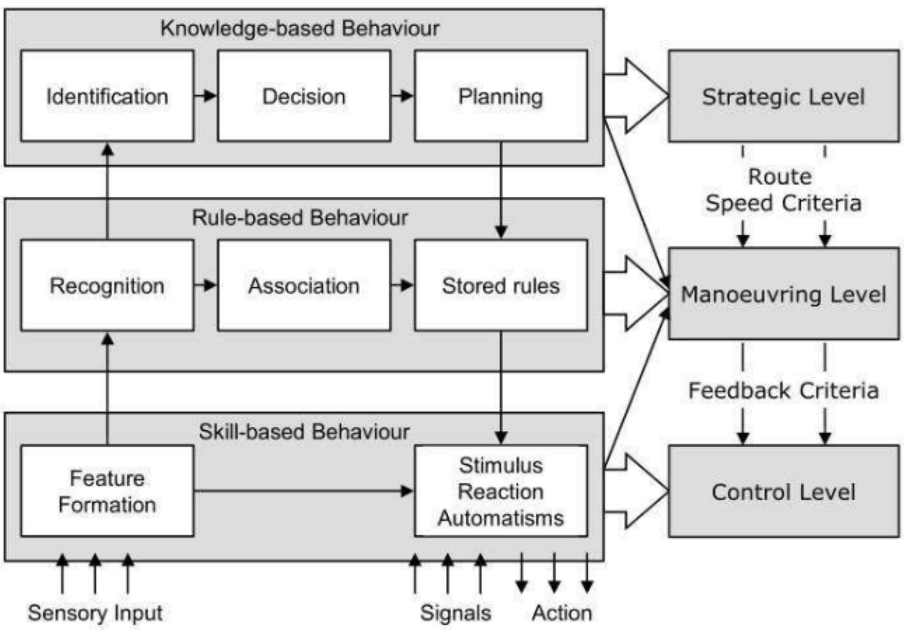


Figure 1.2: Combination of driving task models according to Michon (1985), Rasmussen (1986) and Weller and Schlag (2007)

Here it becomes clear that a vehicle with the current state of the art can handle the skill-based skills as well as the rule-based skills without major problems. However, there are difficulties with the strategic levels of the driving task, since these have a certain degree of uncertainty or complexity that a machine cannot process. The diverse tasks are difficult to implement for an automated vehicle in their entirety, since they involve not only the vehicle's own movement, but also, as described, the interaction with other road users and their intentions in traffic.

In contrast to this, one could state that even a horse carriage is a more autonomous vehicle than traditional cars since horses can dodge obstacles on their own even if the passengers are asleep or busy fending off attackers (Maurer, 2015). That the modern car cannot guarantee this is confirmed by the 1.35 million deaths annually on the road (Haghighi et al., 2020; World Health Organization, 2021).

How does the automated vehicle now become an autonomous agent? The definition of an autonomous agent is close to the definition of artificial intelligence in general. An autonomous agent is any system that perceives its environment and acts autonomously to achieve specific goals and adapt and learn. Compared to artificial intelligence, an autonomous agent has a certain physicality, i.e., the possibility of physically locating itself and interacting with a place or objects (C. Lee & Coughlin, 2015).

A self-driving vehicle meets this definition, as it has to perceive, plan, decide and execute decisive actions. This means that the vehicle must collect all data relevant to the surrounding traffic with the help of sensors or systems to form a representation of its surroundings. These representations include the road surface, all road markings, road signs, traffic lights, and the identification of other road users and their direction of movement. This information can not only be data collected by sensors on the vehicle. It can also be external data such as signals from other road users or GPS satellites. All this data is then collected and fused (Gruyer et al., 2017; Ramos et al., 2017).

With the help of this environmental information, the vehicle has to control its own movement and trajectory planning. However, since traffic is also an interactive and social phenomenon, the car must predict other road users' behavior and plan its actions accordingly. This information about the planned turn-in angle, acceleration, and target speed is then passed to the mechanical components of the vehicle control system so that the vehicle follows the selected path and reaches the desired destination.

The complex interactions between all participants in traffic and the constantly changing environmental conditions result in an infinite number of possible scenarios. So how will enough data ever be collected in testing these vehicles under controlled conditions? Virtual realities can help to cover as many scenarios as possible (Waschl et al., 2019). However, it is essential to distinguish between different levels of simulations: on the one hand, detailed simulations of the environment, and on the other hand, simulations of driving with a cooperating occupant since soon the driver will most likely have to monitor the systems or take control in critical traffic situations. Such simulations in virtual reality are the content of this thesis and will be presented in detail later on. Companies actively researching self-driving vehicles such as Waymo are already using a multiagent simulation environment developed to test algorithms for self-driving cars (Connors et al., 2018).

So far, it is clear that there is an apparent structure within the vehicles that divides the driving task into three separate areas, i.e., perception, planning, and execution. In the following chapter, the functionalities of automated and autonomous vehicles will be explained in more detail to understand the perception and decision-making processes better. However, it is unclear how the vehicle arrives at the conviction to avoid the dachshund on the road from unstructured data such as images from the front camera. Therefore, the way how a self-driving vehicle perceives its environment is discussed below to find a conceptual basis before the legal and ethical issues.

Perception and Decisions of Self-Driving Vehicles

Now that the history of self-driving vehicles in conjunction with self-learning machines has been explored and the driving task has been roughly outlined, this chapter revisits in more detail how goal-directed behavior is generated from the vehicle's sensor capabilities.

To approximate human perception, vehicles need a large number of sensors. A combination of sensor technologies is often used, including cameras, radar, GPS, ultrasound, and Light detecting and ranging (LiDAR) (Kutila et al., 2016; Varghese, Boone, et al., 2015). As with vehicles driven by humans, GPS is primarily used to navigate a map in the range of meters. On the other hand, radar technology refers to the immediate area around the vehicle and allows a rough perception of an object's shape with low resolution, but it can be used to estimate the object's size and speed. However, identification of the object is not possible. Therefore, radar is often supplemented with camera systems. The multiple cameras are often the main source of information for the car, since it usually offer high resolution and allow segmentation, classification and localization of objects in the environment around the car at low cost (Bechtel et al., 2018).

Additionally, the cameras do have a broad viewing range. However, under certain circumstances, cameras have problems fulfilling their tasks, for example, in bright backlighting, insufficient illumination, or heavy fog (Rao et al., 2019) For this case, self-driving vehicles have a LiDAR system.LiDar creates a 3D representation of the environment by emitting and reflecting laser light. LiDAR is similar to radar in terms of perception, but since it is based on light rays, its resolution is so precise that it can be used to classify other road users. Still the signal can be distorted by rain or fog, that is why it is often only one part of the sensor system (Kutila et al., 2016; Yoo et al., 2018). Ultrasound, the last technique in this canon, is used for close-range obstacle detection or in combination with cameras or radar (Mwaffo et al., 2020).

All this sensor technology is, of course, a significant cost factor for such a vehicle, and it is not clear whether LiDAR technology will catch on in a consumer vehicle (Yoo et al., 2018). Therefore, current vehicles on the consumer market do not use LiDAR technology and rely on a more cost-effective combination of radar and cameras (Kutila et al., 2016). Up to now, vision-based algorithms are the most basic and fundamental methods for the detection of the roadside, traffic signs, traffic lights, and other road users (Ouyang et al., 2020). A precise object detection is of utmost importance.

The picture-based perception of objects consists of two processes. One is image classification, and the other is image localization. Objects in the field of view of the vehicle get a semantic label and a position in space. This happens with the help of deep neural networks, in many cases convolutional neural network (CNN). These networks increase complexity in each layer, starting from color or edges and ending at the identification of substantial elements as the intended objects (M. Yang et al., 2019). Without going into detail, CNNs are a powerful tool for reliable image recognition and ultimately enable AI to gain meaningful information out of image data provided by the car's cameras.

With a variety of data representations, the data from the sensors is then used to make predictions about possible movements of other road users in the near environment of the car (Bechtel et al., 2018; Ramos et al., 2017; Rao et al., 2019). Based on these prediction and interpretations, the car then plans and execute actuate maneuvers. It is important to note here, that the car needs to consider that not every signal of the sensors is true and that human road users are only hard to predict. Just like we saw in figure 1.2 the car has to plan its route, navigate itself along this route while considering the movement of other road users and control its movements during the transition.

1.3 | Classifications of Autonomy Levels

To understand the difference between autonomous driving vehicles autonomous driving vehicles (ADVs) and semiautonomous driving vehicles, it is necessary to understand the general definition of autonomous and semiautonomous machines since the differentiation between a highly automated and a fully automated vehicle is not so clear as it seems. The three competing categorizations are taxonomies of the society of automobile engineers (SAE), the Bundesanstalt für Straßenwesen (BASt), and the national highway traffic safety administration (NHTSA). In the further course of this thesis, however, only the SAE taxonomy will be discussed in detail since this categorization has become established in international research.

By definition, the fully autonomous system decides without human intervention of any form (Shadrin & Ivanova, 2019). Accordingly, an autonomous system would be able to act on its own in response to unknown and unexpected traffic events. However, before a fully autonomous vehicle is achieved, there are preliminary stages where driving tasks are automated while still relying on the human driver as a supervisor. This is the difference between highly automated driving, which only works under certain conditions, and the fully autonomous vehicle. Therefore, automated systems are less autonomous than fully autonomous systems (Norris & Patterson, 2019).

SAE has created a classification of automated driving functions in six levels. This classification is very technical, and especially the higher levels of automation are indistinguishable for nonprofessionals (Shadrin & Ivanova, 2019). For the technical description and, later in the thesis, the ethical and legal classification, the SAE taxonomy is the most appropriate. As shown in figure 1.3, automation ascends from level 0, no driving assistance to level 5, the fully autonomous vehicle. To classify an automated vehicle, the question must be asked to what extent the humans or the machine is responsible for the dynamic driving task (DDT). This is defined as "the set of real-time operational and tactical functions required to operate a vehicle on the road, excluding strategic functions such as trip planning and destination and waypoint selection." (Shadrin & Ivanova, 2019).

It is precisely this division of the DDT that then determines the classification of the automation level, which will now be examined in more detail.

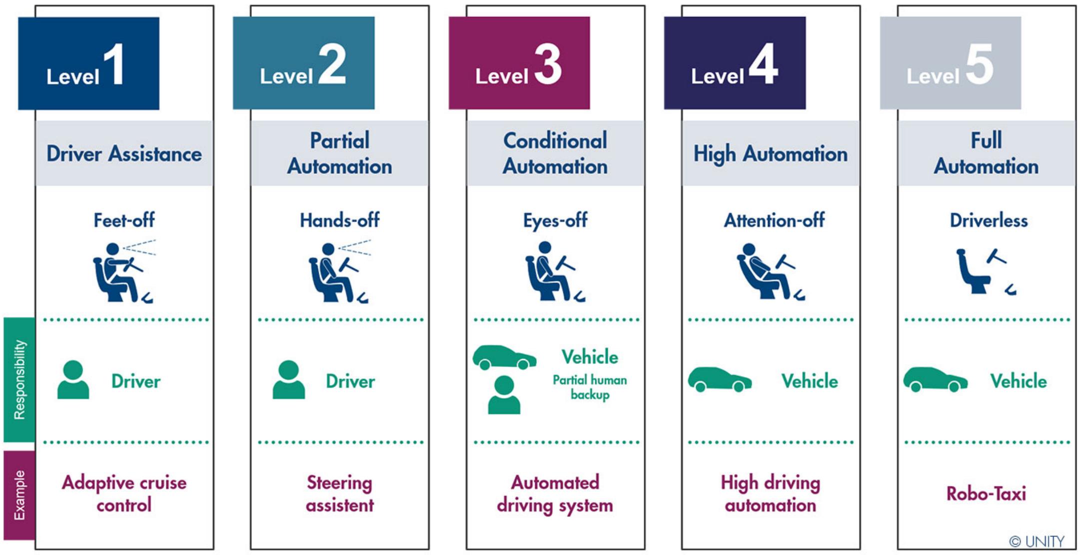


Figure 1.3: Overview of SAE levels of automation
(Gaupp & Katzenbach, 2019)

Level 0: No Driving Automation

This level of automation accounts for a large number of used vehicles on the road today and means that the DDT is handled exclusively by the human driver. Even if the human driver is supported by active safety systems like a lane departure warning system or an emergency brake assistant, these features do not count toward fulfilling the DDT as they do not technically control the vehicle but only support the driver.

Level 1: Driver Assistance

At this level, the human still has tactical control of the vehicle in real time. However, a single assistance system is used for the DDT sub-task of either the longitudinal or lateral motion control. The assistance can be, for example, an adaptive cruise control system that measures the distance to the vehicle in front and maintains a predefined distance. Here the human decides when the driving automation feature is appropriate and takes over the entire DDT when desired or required. Even with an activated system, the remaining tasks have to be performed by the driver at all times. Nearly all new cars are fitted with at least one system that controls sub-tasks of the DDT.

Level 2: Partial Driving Automation

Level 2 is defined by automated lateral and longitudinal vehicle motion control that disengages when the human driver engages in the DDT. With longitudinal and lateral motion control, the driving automation system controls a large amount of the DDT. The system is often enabled by many sensors such as radar, cameras, and ultrasonic sensors. Level 2 includes active road user detection, lane-keeping systems that actively direct steering, traffic sign recognition and speed control, and emergency braking. However, while the vehicle controls the actions in traffic, the human driver is still responsible for supervising the system as well as the object and event detection and response (OEDR), that is a subtask of the DDT. Like in the first level, the human driver engages and disengages the driving automation systems as desired or required. A prominent example here is the Tesla Autopilot.

Level 3: Conditional Driving Automation

This category is defined by an entirely automated DDT and OEDR under certain conditions. It means that the car can fully control itself in a specific operational design domain (ODD), meaning the environment it can operate in. Nevertheless, this still implies that the human driver should take over control when the car demands so in a take-over request (TOR). This fallback to the human requires the car to estimate when an ODD exceeds the car's capabilities and disengage within an appropriate takeover. A Fallback could happen when a failure in the sensor systems occurs or the car faces uncertainties due to an unknown environment. When the Fallback ready user does not respond, the vehicle has to use a failure mitigation strategy. One example for this could be coming to a complete halt right in place. Also, the human driver is still able to override the car's decisions at any given point of the route. Within the ODD, the driver does not have to engage in the DDT nor the OEDR. One practical example of this is the Audi Traffic Jam Pilot (Audi AG, 2021; Blackman, 2018), where the car fully controls the vehicle motion only on highways at up to 60km/h.

Level 4: High Driving Automation

In technical terms, the progression from level 3 to level 4 is a big step since the human fallback is eliminated. Like in Level 3, the system is responsible for the DDT and OEDR under certain driving conditions, but it does not expect the human driver to intervene while the automated driving system is engaged. In this case, the driver becomes a passenger. Ergo, the car must deal with system failures and unexpected traffic events in the ODD but may ask for a takeover when faced with an ODD limit. Since the human driver is not expected to take control within the ODD, the car falls back into a minimal risk condition. These minimal risk conditions represents a halt of the vehicle. Still, the driver can request a disengagement of the automated system when desired. However, the use of such vehicles is currently limited by existing infrastructure and legislation. Currently, it is not clear who bears responsibility in the event of an accident as the human driver has no active part in decision-making in highly automated driving conditions. This uncertainty results in these vehicles being restricted to limited areas suitable for driving a predetermined route. One example of level 4 automation is the Hub Chain collaborative project between the Deutsches Zentrum für Luft- und Raumfahrt (DLR) and Stadtwerke Osnabrück (van Tongern, 2018). The project involved the development of the autonomously driving bus shuttle "Hubi," which is qualified as an SAE level 4 vehicle and operates (within the ODD) without a human driver on the ICO Campus Westerberg of the University of Osnabrück.

Level 5: Full Driving Automation

In contrast to all previous levels, no human attention or supervision is required for the DDT or DDT fallback in any driving conditions (not ODD-specific). The system plans and executes all necessary functions without a possible fallback at any time and under unlimited conditions. This means that all humans in the vehicle are classified as passengers; consequently, the system does not need to take control and therefore does not need a driver. Ergo, the human involvement in the DDT is completely eliminated since the car is able to handle fallback situations with minimal risk when system failures occur. The system only disengages when the car reaches a minimal risk condition or the human passenger requests to takeover. But it seems unlikely that the human will be able to manipulate the car in real time since these vehicles will most likely not be equipped with a steering wheel or pedals.

As noted, the SAE classification is an international standard that does not specify certain methods or techniques but rather levels of automation. This taxonomy can be assigned permanently and remains unambiguous and compatible with the simplified model of user communication presented here. The simple terminology proposed above describes driver roles and remains consistent with the previous work of the BASt, from which the described classification differs only in highly automated and autonomous driving. In summary up to level 2, one can speak of assisted driving. Level 3 automates all functions of the driving task under certain conditions so that the vehicle can temporarily drive itself. These functions make it possible for the first time to perform a non-driving-related task (NDRT) safely while driving. The use of this technical option is legally permitted in Germany, the law on automated driving being passed in 2017. From level 4 onward, vehicles can drive autonomously under certain conditions without drivers having to take over -in other words, the functions autonomously perform the driving task in pre-mapped and highly controlled areas.

The SAE automation levels are critical in developing such systems and as a basis for legal assessment since these levels assign responsibility and roles to the human driver or passenger and the automation system. However, self-driving vehicles are a complex legal issue. Since regulation of this technology can be a potential threat, or enabler, for self-driving vehicles, legal problems will be described in the next chapter. However the following chapter remains without presenting great detail or possible solutions as this would far exceed the scope of this thesis.

Legislation of Highly Automated and Autonomous Driving

As already indicated, there are substantial legal difficulties in applying artificial intelligence due to the problem of the inexplainability of decisions and the residual uncertainty of such systems. Inexplainability and uncertainty will now be considered in the case of self-driving vehicles.

Autonomous driving raises complex questions regarding liability in the event of an accident or error. It is of the utmost importance, however, to solve these problems in order to realize the technology (Lohmann, 2016). In principle, there is agreement in the German government that the technology will most likely contribute to improving the safety of individual traffic and should therefore be actively supported. Nevertheless, SAE levels 4 and 5 cannot be achieved in accordance with the current legal regulations (Dix et al., 2021; Greger, 2018; Lutz, 2020). It is still not clear who or what is responsible in the case of a system failure or a crash scenario when the automation system is engaged and the human driver is disengaged (Gurney, 2013).

The technology raises the question of who is liable if, for example, a driver suffers a stroke while driving and the vehicle controlled by assistance systems causes damage. In a contrived extreme case, a driver suffers a stroke while the autonomous systems are engaged. The level 0 vehicle with the incapacitated driver would have crashed in a field. In our case, the assistance system stays on track and runs over a group of pedestrians. Is anyone responsible for this situation, or could this be labeled a risk of life?

This question is addressed because of the Vienna Convention on Road Traffic of 1968, according to which every vehicle needs a human driver who has full responsibility for the vehicle (IT Commitee of the United Nations, 1968). However, the Vienna Convention was amended by the United Nations in 2014. Now, not only humans are permitted as drivers, but also systems with which a passenger car drives autonomously if the driver can stop them at any time in an emergency (Bundesministerium für Digitales und Verkehr, 2017). Still, the driver remains liable for every action of the vehicle even if it is driving autonomously. While vehicles in levels 1, 2, and 3 can be accommodated within this legal framework, this becomes complicated for levels 4 and 5. Therefore, these vehicles are currently not legally compliant (Greger, 2018). This example shows that the fundamental change in technology makes adjustments to the legal framework imperative – throughout the EU.

A revised law has now been in place in Germany for a few years. In June 2017, the legislature enacted the new § 1a of the Road Traffic Act (StVO) (Bundesministeriums der Justiz und für Verbraucherschutz, 2013) to allow the operation of a motor vehicle with higher driving automation if the function is used as specified. However, the original problem of responsibility is not directly addressed here because the vehicle driver remains responsible even if this person does not drive the vehicle. According to § 1b StVO, the person in the driver's seat may turn their attention from road traffic but must remain "perceptive" so that the control can be taken over at any time. However, there is no case law on this subject yet (Greger, 2018). This means that it has not yet been clarified how a person behind the wheel is liable to prosecution if they are not driving. Since the German legal system is linked to an action or an omission in terms of criminal liability, the software, hardware, and the manufacturers of the vehicles are not considered because they are not actively involved in road traffic (Greger, 2018; Lohmann, 2016).

When using vehicles of level 3 and higher, the driver will retain responsibility even if he or she is not actively involved in controlling the vehicle. Potential customers must be confident that the system will operate with complete reliability and that the person in the driver's seat will be allowed to devote his or her time to tasks other than monitoring the system. This trend toward more and more automation cannot be reversed, and it can be assumed that more and more parts of the DDT and OEDR will be transferred from the driver to the system in the coming years. However, this still means a lot of open questions for the justice system. As an acting system, cars deciding over life and death will be breaking new ground (Coelingh & Ekmark, 2019).

The question of liability in accident situations is closely linked to the question of decision-making. Who or what is liable in the event of property damage or personal injury? How should a self-driving vehicle should react in a critical traffic situation in the first place? As we have seen, this is a minor issue for the self-driving vehicles of levels 0 and 1 but already becomes substantial in level 2, as seen in various accidents where drivers were inattentive during activated automation (Penmetsa et al., 2021). Although there is no legal basis for the higher levels 4 and 5 (Lohmann, 2016), these incidents call for new ways in the human-machine interaction to enable collaborative decision-making for fast and reliable reactions (Roche et al., 2019).

1.4 | New challenges in Human-Machine Interaction

In the previous chapters, we looked closely at the basics of perception in self-driving vehicles and how they are technically defined and divided into automation levels. We have also seen that several open legal problems need to be solved. However, this is not the end of the long list of open problems. Another challenge for the realization of highly automated and autonomous vehicles is, apart from all remaining technical and legal issues, the problem of missing consumer acceptance (C. Lee et al., 2019). If this is nonexistent or found only among individual interest groups, the technology is on the brink of failure. Only if there is broad acceptance among the population and the technology is seen as safe and valuable by potential customers will it be possible to scale features of self-driving vehicles in individual traffic (Nees, 2016).

Therefore this chapter covers the basic idea of general acceptance as a critical factor for realizing highly automated and autonomous vehicles.This chapter also covers the topic of social implications of this technology, or how our cities and lives will possibly change due to the emergence of fully self-driving cars (Yaqoob et al., 2019). Apart from the question of fundamental acceptance and the question if we, as a society, want to use self-driving cars on our roads, there is also the question of the human-machine interaction inside the car (Wachenfeld et al., 2015). Since this is the main topic of this thesis, a substantial subsection concerns an experiment on human supervision in highly automated vehicles and the problem of the out-of-the-loop unfamiliarity (OOTLU)

Acceptance and Social Implications of Autonomous Cars

As already covered, the realization of ADVs is bound to the acceptance of autonomous machines. Different survey reports about customer attitudes toward ADVs state that acceptance seems to be critical for realizing this technology, but also that much doubt exists concerning safety and privacy (Bergmann et al., 2018; Gillmore & Tenhundfeld, 2020; Nastjuk et al., 2020; Raue et al., 2019). These concerns might be irrational since the risk of a crash in an ADV is thought to be significantly lower than in a human-controlled car (Yadav & Velaga, 2020).

The logical question here is why this technology is attractive at all and if the possible outcome is worth the current efforts. As already mentioned several times, in addition to the possibility of fewer accidents and traffic fatalities, there is the better climate balance of self-driving vehicles (Chehri & Mouftah, 2019). The emission reduction is because automated and autonomous vehicles drive slower and more consistently than the average human driver (H. Li et al., 2020) and because the introduction of fully automated vehicles will likely change our entire mobility behavior (S. Pan et al., 2021). Because of the status quo, an individual vehicle is an object that takes up mostly public space for an average of 23 hours a day. With the introduction of robot cabs, this could fundamentally change (Silva et al., 2021). It is conceivable that there will be not, or only rarely be privately owned vehicles, while on-demand service will be much more likely (Cusumano, 2020). The on-demand vehicles will be used constantly, eliminating the need for large parking spaces. The space thus freed up by reduced parking spaces, roadways, and other infrastructural buildings such as gas stations can thus be accessible to the public again. Therefore, the self-driving car can substantially contribute to a green vision of sustainable inner cities (Campisi et al., 2021; Chehri & Mouftah, 2019).

Additionally, there is hard evidence that this technology will have a powerful and positive impact on how we use roads (Fagnant & Kockelman, 2015; Krueger et al., 2016; S. Pan et al., 2021; Ryan, 2020). As mentioned before, the roughly 1,200,000 people dying in traffic globally each year could be decimated (Hevelke & Nida-Rümelin, 2015a, 2015c; World Health Organization, 2021). Also the total emission of toxic exhaust gases would be significantly lowered through fuel-efficient driving and avoidance of congestion even if future ADVs still run on combustion fuels (Chehri & Mouftah, 2019). Also, pople with limited access to individual traffic would be included. Physically impaired and even blind people could participate freely in individual traffic (J. Yang & Coughlin, 2014).

The major premise for the self-driving car is that it should take away psychological stress while increasing road safety (Koo et al., 2015; Reimer, 2014). Various authors point out the importance of perceived risks on trust (Gillath et al., 2021; J. D. Lee & Kolodge, 2020; Raue et al., 2019). Perceived risk is a substantial factor linked to trust, particularly concerning the decision to use an automated device or not (Gillath et al., 2021; J.-H. Lee & Song, 2013). Their results confirm the influence of trust on automation when trust is considered both as a direct determinant of behavioral intention and as an indirect influence through perceived usefulness and perceived risk. In their publications, trust reduces the perceived risk depending on the expected probability of a negative situation. When drivers trust ADVs, they assume that the vehicles will behave as expected, reducing the perceived risk of a negative situation. The term trust is therefore defined as the users' attitude that the vehicle will act according to the human objectives in any given situation (Stephan, 2019). Acceptance in this context describes a broader term that refers to consenting to the usage of a self-driving vehicle.

We will see later that the perceived risk and the intention to use such a vehicle are closely connected to acceptance rather than trust. Additionally, it will be shown that people with little technical knowledge have stronger reservations toward self-driving cars. We suggest that people cannot form a mental model of the car and how it is perceiving and evaluating its environment. Because these participants cannot formulate models of how a system is working, it becomes generally unknown in its core function - resulting in skepticism. Despite the doubt, automation's crucial point is introducing abilities related to human-centered activities that go beyond human capabilities. However, this development also poses new problems, which will be discussed in the following section.

Human Supervisors and the Out-of-the-Loop Problem

This section deals with a completely different problem that comes from the human user. Even if it were possible to increase acceptance within society immediately without problems, user competence plays a crucial role in partially and highly automated vehicles. The problem here is the role of the only partially intervening driver, who is in principle dedicated to tasks other than driving but is faced with special requirements precisely when they are inattentive: the takeover request.

With increasing automation, it will be possible that human drivers turn their attention from driving-related tasks toward car-unrelated tasks, like scrolling through the smartphone or attending video conferences. During this time, the car takes over the DDT and the OEDR. But it is necessary to take back control as fast as possible once the self-driving car requests a fallback (Shalev-Shwartz et al., 2018).

Here, it becomes essential that the human can monitor the current traffic environment without paying attention to it. Research could show, that mere system awareness is not as effective as tailored visual and auditory cues about objects in front of the car (Wintersberger et al., 2020). Therefore, a human-centered design for a fast, precise, and focussing monitor system is needed (Roche et al., 2019). The interaction between human users and responding machines has been the focus of scientific studies for many decades, now starting in the late ’60s (Dietsche & Kuhlgatz, 2014). In the context of self-driving vehicles or care robots, this means that the human-computer interaction becomes more of a human-machine cooperation (Weyer et al., 2015).

In cases where the car takes over the traditional tasks of a human driver, it becomes an agent that needs to make decisions and interact with the human, as well as other road users. This double role transforms the driving experience of a human driver, using the car as a tool, to a collaborative activity, where agency is exchangeable between the human driver and the machine (Norman, 2015; Weyer et al., 2015). This exchange needs a form of direct interaction. A potential starting point could be human-directed communication in a verbal interaction like the Siri system introduced by Apple or the SYNC system by Ford since these systems are accepted by customers already (Tulshan & Dhage, 2018). A study at Colorado State University showed that the acceptance of a novel technology could be increased if said technology possesses a name, a voice, and a gender and, most importantly, if it can report about itself and its functions. The authors also suggest that anthropomorphic features in a car could be a promising approach since car owners often treat their cars as autonomous creatures already (Waytz et al., 2014). It might bring back a breeze of the fascination of mass motorization between the 1920s and 1950s (Kröger, 2015). While a fundamental form of verbal communication is already feasible in modern vehicles, the question of task division and cooperation is still to come since this requires that the car be able to take over a substantial part of the DDT.

Driving a car requires a particular set of skills, like using the gearbox, controlling the car with a steering wheel, or evaluating potentially risky situations and the behavior to control these situations. However, in the potential future, the human will no longer be a physically active part of the vehicle in a fully automated ADV, resulting in a possible loss of these competencies (Bainbridge, 1983; Kröger, 2015). Humans will ultimately be rather more like a supervisor, with new tasks and responsibilities (Endsley & Kiris, 1995; Engström & Hollnagel, 2007). Of course, there is a need for a new understanding of the technical aspect of ADVs. As a supervisor with the responsibility to keep a highly automated car in a safe condition, the passenger needs to understand the essential functions of the car, at least, be able to identify possible sources of failure, and acquire the knowledge and skill to keep the car in a safe condition in case the ADV is not able to drive itself anymore. A risk here is that users will slowly forget how to drive a car independently (Norman, 2015). This risk dramatically increases with more automation. Users slowly rely on the car systems to work ideally — and in most cases, they will. However, there is still the possibility of a system failure. What happens if the responsible driver that has not touched the steering wheel in quite some time is forced to react immediately (Bainbridge, 1983; Endsley & Kiris, 1995; Norman, 2015)? This means that users unable to drive on their own maybe find themselves in a situation where the car demands human control, for example after a crucial system or sensor failure, but the human driver is not able to assess and react to the situation anymore (Bainbridge, 1983; Norman, 2015).

One consequence could be that the further development of ADVs defines new roles for the passengers. This means that the functions and actions of the car have to meet the humans' capability of information processing and their needs and expectations (Hallerbach et al., 2017; Weyer et al., 2015). Although cars are not on the technical level to take over the DDT and OEDR yet, there are examples of mobility where more advanced automation is already standard; for instance, modern high-speed trains and airplanes (Chialastri, 2012; Spring et al., 2009). Of course, this comparison is imperfect. A strict set of rules defines air and rail traffic, and these systems divide tasks among a group of responsible entities. Also, pilots are well-trained experts able to supervise their system in nearly every technical aspect (Chialastri, 2012). Most importantly, however, a pilot is often faced with problems in a large space: the airspace. Logically one can assume that, as long as no wing or the hull of the airplane is critically damaged, the pilot has several minutes to assess a problem, while being assisted with professionals in the ground control tower (Chialastri, 2012). In contrast, road traffic is a system without such strictly defined rules. Sometimes driving over a sidewalk to avoid an obstacle or a possible collision is an excellent thing to do that a self-driving car would not. For example, humans might drive over the sidewalk to make room for an ambulance, whereas the air and rail space has strictly defined protocols for comparable situations. This example highlights the importance of a human supervisor. This, this supervisor has the unconditional task of quickly and precisely understanding the context of the current driving situation and executing actions that would render the rules of an automated vehicle absurd.

This comparison supports the argument that it is insufficient to optimize autonomous systems independently but crucial to also focus on the human-machine interaction when human operators function as system supervisors responsible for correcting system failures or limits (Bengler et al., 2020).

Further development of self-driving cars could lead to three different scenarios: overrated trust in faulty automation, the loss of cognitive or manual skills, and the loss of the capability to evaluate potentially risky situations, resulting in inappropriate human behavior. These phenomena are described by the out-of-the-loop unfamiliarity (Bainbridge, 1983; Endsley & Kiris, 1995). A proven way to handle this issue is to keep people in the loop by passing ultimate responsibility into human hands regardless of automation, just like in aviation.

However, as examples from aviation show, pure supervision as a task is not enough since humans are unsuited for prolonged attention, and it is questionable whether the cognitive load is reduced (Chialastri, 2012; Spring et al., 2009). Humans are not well suited for long periods of vigilance: Tests show that during prolonged periods of supervision, pilots tend to be out of the loop, resulting in them not being able to assess information quickly enough in a possibly dangerous situation (Endsley & Kiris, 1995; Engström & Hollnagel, 2007). Therefore, it is necessary to find an alternative that, despite higher automation, allows humans as supervisors to receive enough information to make informed decisions even if they are not attentively following traffic events.

If we now assume future automation will take care of certain events in traffic like it is anticipated by the SAE, these events includes other drivers, or possibly tricky road conditions and the current state of all of the car's sub-functions. To hold a human in the control loop, one approach is to inform the inactive passenger about the vehicles state to enable the passenger to take over control in a short amount of time. This information has to be represented in a manner that is natural and does not require continuous attention (Norman, 2015). Too many signals are worse than not enough since this possibly results in distraction; ergo, the human driver is out-of-the-loop (Norman, 1990; Roche et al., 2019; Spring et al., 2009). Examples of systems, giving information about the current driving state include almost all lane and distance assistance systems, which activate a pulse in the driver's seat or even steer actively to keep a car in a lane. However, these seem to be somewhat abstract representations of the current driving status, and apparently, these representations are a possible source of stress in passengers since they are hard to interpret (Naujoks et al., 2016).

Quickly, the question arises: What is needed to avoid stress reactions and build human trust in a way that potentially keeps human supervisors in the loop? In the following, mental models will be introduced that appear to be an appropriate theory for building trust in ADVs and emphasize a human-centered interaction. Jay Wright Forrester introduced this theory in 1971 for educational psychology, and it describes how humans process functionalities in the world (Forrester, 1971). These mental models are cognitive-emotional representations of objects, object relations, and processes. As known in cognitive psychology, human beings develop simplified models of the functions and processes of their environment (Johnson-Laird, 1986). These models are used to gain orientation, as well as to understand and predict certain events. These are dynamic processes that three features can describe:

Mental models are processed in the working memory and enable the simulation of possible actions. Thinking about possible other plans or outcomes means altering mental models. The second feature is that mental models can represent causes and relations between events in an abstract form. Therefore, they generate a causal understanding of how a system works. Lastly, these models are also externally changeable through experiences; therefore they are linked to learning. If an outcome of a repeated action differs, it will change the mental models (Johnson-Laird, 2006).

Hence, mental models are based on context-specific expectations and the user's knowledge of the system. This also applies in the context of highly autonomous driving situations. Nevertheless, the decision regarding the human use and acceptance of technology is not based solely on rational understanding alone because attitudes are not entirely changeable through information, as we will see in the following chapters. More likely, new information is processed selectively to harmonize with already existing models (Johnson-Laird, 2006). Tests of one possibility to increase human acceptance of technology will be described in this thesis: different feedback and feedforward strategies used by an anthropomorphically acting ADV for crucial safety-related situations in traffic.

However, how should these systems be tested if the highly or fully automated level 3, 4, and 5 vehicles needed for research do not yet exist? The proposed approach is immersive virtual reality virtual reality (VR). This offers the possibility to create counterfactual environments and technologies and thus to generate natural behavior under controlled conditions. How this works exactly is explained in more detail in the next section.

1.5 | Virtual Reality Simulations and Human Behavior

The technology of virtual reality has been steadily growing over the past years. It lets users experience three-dimensional computer-generated environments via the use of a head-mounted display (HMD). The HMD combined with the three-dimensional computer environment enables the user to experience a dynamic control of the viewpoint in a responsive virtual world (Steuer, 1992). While this might seem limited at first, this enables a feeling of presence and immersion within these virtual worlds that is inherently different from two-dimensional experiences of virtual worlds in classical screen setups. The feeling of being physically present in a virtual world distinguishes VR from all preceding technologies (Slater & Wilbur, 1997).

In general, VR allows the presentation of potentially dangerous situations in a safe and controlled environment while the researcher records behavioral information of the subject. This information includes the subject's head, hand and eye movements, as discussed later in this chapter. Additionally, the creation of virtual worlds represents a high degree of design freedom. It is possible to represent self-driving vehicles in realistic scenarios, although this technology does not exist yet. It is also possible to give the participant a realistic impression of critical traffic situations, which any ethics committee would reject in the real world. Here edge cases and extremes can be tested. Likewise, it is possible to experimentally represent the design and functionality of vehicles since the virtual environment is not bound by road laws or limitations in manufacturing.

This is where everything necessary for research into human-machine interaction in the context of self-driving vehicles comes together. The problems of decision-making, cooperation with the human driver, and the OOTLU known from the previous sections can thus be addressed. Nevertheless, this is not limited to the driver's interaction with the vehicle because tests of highly automated vehicles in classic test setups have only a minimal yield of possible test scenarios. In principle, there is probably an infinite number of possible traffic situations. Of course, this cannot be covered by test scenarios in controlled physical test environments. New methods are needed for a large number of complex situations under a wide variety of conditions. Virtual reality can make a substantial contribution to the training of self-driving features in modern cars.

For this purpose, too, we would like to provide two toolboxes that can be customized as desired. For such an environment to offer real added value, it should represent a realistic representation of the physical behavior of the environment and other road users. Likewise, sensors' potential properties and limitations, e.g., virtual camera systems, should be included to enable adequate testing. In order to implement a multitude of possible conditions and tests, which allow testing in different automation levels and with freely configurable virtual components, we have developed the project Westdrive. We put much effort into making the world as realistic as technically feasible, and we primarily focused on including traffic events as modular adaptive components, which will be part of the later chapters. However, before moving on to the first studies, we take a closer look at how virtual reality devices work.

Virtual Reality

Our biological eyes have two different angles on a sighted object in the real world. This indicated by a slightly different projection of fixated objects on the retina. This, in turn, means that both eyes perceive a different change in visual angles as shown in figure 1.4. Our brain can detect these small differences and form an impression of them by perceiving size and distance. However, much of the visual input from our two eyes is the same. Thus, it is possible to combine both images (Haber, 1978; Holmqvist et al., 2012). This convergence is also mathematically possible, which in turn enables the development of stereoscopic displays. Splitting content into two stereoscopic, two-dimensional images displayed separately for each eye is the basis for all modern virtual reality headsets to immerse their users in a three-dimensional environment (Clay et al., 2019; Duchowski, 2017).

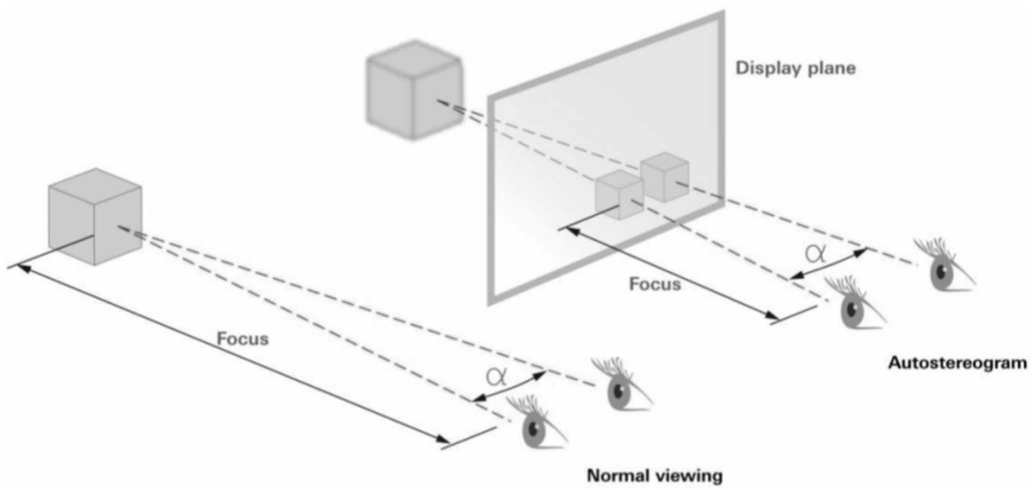


Figure 1.4: Depiction of the HMD display plane compared to stereoscopic vision (Anwar, 2019)

However, the question arises about how the user's natural movements can be transferred from the real environment to the virtual environment. The transfer is achieved by recording the movement of the devices guided by the user. There are currently two general approaches to this. The first is the so-called inside-out tracking, where cameras on the outside of the HMD are used to orient the device in the real-world space. In this case, the calculations of rotation and orientation depend only on the device itself. This method is mainly used by Oculus (Facebook Technologies LLC, 2021). This form of tracking does not require any external devices, and the user can move freely. The second approach is called outside-in tracking, where cameras detect a passive device. One subform of this is lighthouse tracking, which needs additional hardware to locate the VR devices in space. This approach is used by HTC and Valve, as well as by Sony for the Playstation VR (Sony Interactive Entertainment LLC, 2021; Valve Corporation, 2021). In lighthouse tracking, the position of the HMD is calculated based on device position and angular speed. Here infrared light from external emitters is used to create a grid invisible to the human eye to enable tracking 1.5. It has the benefit of precision because the tracking features are not dependent on environmental factors such as environmental illumination (Gourlay & Held, 2017).

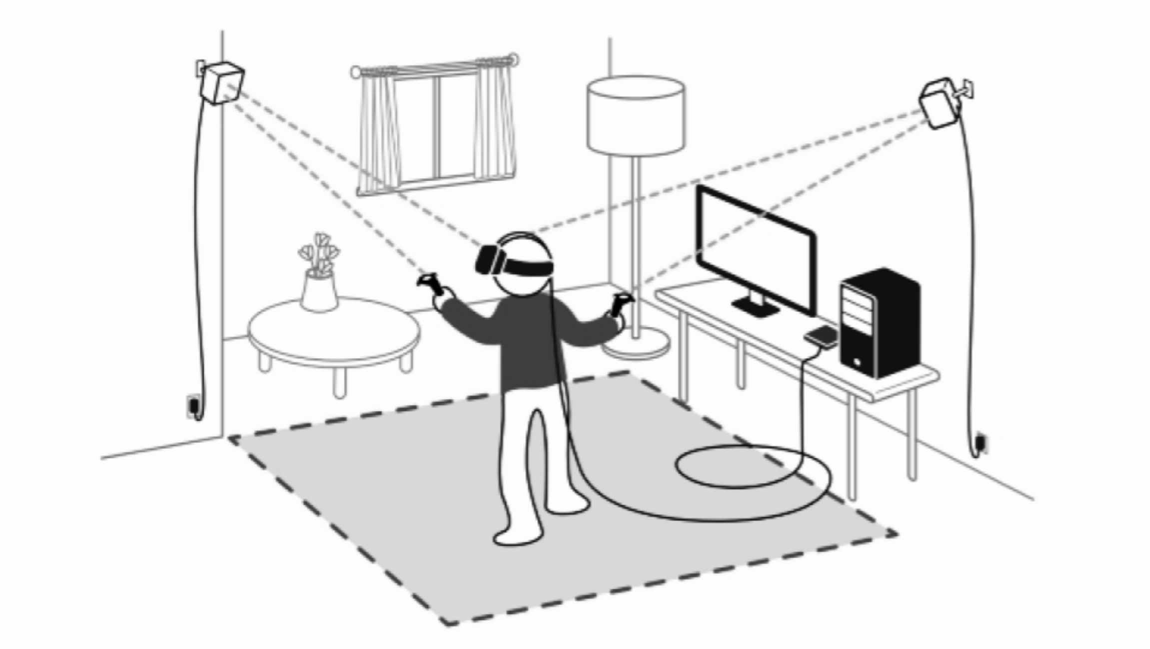


Figure 1.5: Overview of the lighthouse or base station tracking (Yuan, 2021)

With both approaches on tracking the user's movements, however, the user's freedom of movement is limited. Be it limited, be it by the range of the lighthouses or the possibility of the HMD to find its way in less suitable, for example, brightly lit and reflective, environments. Another problem is motion- or cybersickness. This form of sickness occurs when there are differences in motion perception and the device's representation of motion. While it causes a feeling of discomfort, it completely breaks the immersion of the virtual environment, which is especially important for scientific experiments. The causes of cybersickness include a lack of appropriate tracking and a limitation in representing the sense of balance (LaViola, 2000). A ride in a virtual roller coaster does not feel the same as in real life since the sensory input from the vestibular and proprioceptive systems is missing. The difference between the visual sensory input and the missing proprioceptive and vestibular input can cause dizziness and nausea. Nonetheless, these issues are constantly being improved, and new ways to deal with them are developing steadily (LaViola, 2000; Veličković & Milovanović, 2021).

In this thesis, studies in virtual reality are presented. In these experiments the HTC Vive is chosen as the HMD since, in our view, it is the HMD with the best price-performance ratio. The reason is an excellent tracking performance, a large field of view, and an acceptable resolution. Also, this device can be combined with external eye tracking devices and is delivered with a built-in eye tracker by the manufacturer (Ahmad, 2020; Valve Corporation, 2021). The two handheld controllers and headset of the Vive use 70 combined sensors to calibrate the positions of the controller and headset, measuring the time difference in sending and receiving the emitted signal (Ahmad, 2020). The HTC Vive originates from the Valve Cooperation, an online game distribution company that entered the VR market in 2016. An account at the online gaming platform Steam is necessary to use the HTC Vive and the HTC setup software (Valve Corporation, 2021).

With the help of these devices, in conjunction with a computer and 3D programs, such as game engines like Unreal or Unity (Epic Games Inc, 2021; Unity Technologies, 2021), it is possible to create virtual and counterfactual worlds. VR already sees use in a lot of different areas. In medicine, it is becoming more common to use VR to get experience in areas such as surgery, where it is usually hard to get hands-on experience as mistakes can endanger a patient's life. Furthermore, it has been shown that skill acquired this way does transfer over to real-life situations to some extent (Butt et al., 2018). Another area is acquiring driving skills, not only for car drivers but also for pilots. For psychotherapy, VR is being used in the context of exposure therapy since it allows for confrontation with the object of fear inside a controlled and safe environment compared to more stressful real-life situations (Riva, 2003). Therefore the virtual environment is a possibility to examine and understand phenomenons that would be inaccessible through research in the real world (Anthes et al., 2016; Butt et al., 2018; Z. Pan et al., 2006).

Human Behaviour in Virtual Worlds

Especially in science, VR has several advantages. The possibilities are nearly endless; the only requirements are sufficient computational power for the high-resolution output to two displays and sufficient technical know-how to create a virtual environment. It allows researchers to create precisely the controlled conditions of interest while providing most of the advantages of a classical laboratory experiment. Especially for complex environments close to real-world scenarios, it is of benefit that the virtual environment does not change between participants, and everything inside of VR can be controlled to a point where random occurrences and disturbance factors can be nearly eliminated (Parsons, 2015). This way, it offers a standardized, controllable research environment that provides almost unlimited possibilities for experimental setups. Combined with the possibility to provide open access to experimental code, reproducing even the most complex studies will become possible under stable conditions in VR (Wattanakriengkrai et al., 2020).

To additionally benefit of control and reproducibility, it is possible to combine VR with other methods such as eye tracking (Clay et al., 2019; Holmqvist et al., 2012). The eye tracking technique combines camera recordings of the eye with image processing to calculate the position of the center of the pupil and the corneal reflection (Duchowski, 2017) and therefore offers the so-called point of regard (POR) (Munn & Pelz, 2008). It enables an analysis of where a subject is looking, whether the eye got there through eye rotation or head motion. In terms of applications, eye tracking is present from neuroscience and psychology to industrial engineering and from human factors to marketing, advertising, and computer science (Burke, 2018; Riva, 2005; Wolfartsberger, 2019). Although VR eye tracking has not yet reached the same level of performance as classical eye tracking devices (Ehinger et al., 2019), it is nonetheless usable for scientific research and is a focus of this thesis.

The the idea of combining VR and eye tracking is not a recent approach. It had already been thought of nearly two decades ago (Hua et al., 2006). However, as described by Clay et al., 2019, this technology has recently been taking off due to the rapid advancement in VR devices in the consumer market, combined with the continuous development of computer hardware, making it affordable and highly accessible even to the regular consumer market (Anthes et al., 2016). Furthermore, as has been shown, given the right circumstances, eye tracking data measured in VR is good enough for fundamental research to be carried out (Holmqvist et al., 2012).

When designing a VR experiment in combination with eye tracking, it is important to keep in mind that the experimental setup can have an enormous impact on the presentation and experimental data. First of all, displaying two high-resolution screens in the HMD is linked to a high computational load that exceeds that of screen-based experiments. It affects the displayed frames per second in an experiment. A frame rate that drops below 30 frames per second breaks the immersion since movements are experienced as less natural. It can also lead to cybersickness in the participants. Since eye tracking itself needs additional computational power, working with VR in combination with eye tracking must always be designed such that motion sickness is diminished as far as possible, not only for the wellbeing of the subject but also for the most unbiased data acquisition. It is worth critically checking the flexibility of the environment, freedom of subject movement, and the number of possible interactions (Clay et al., 2019; LaViola, 2000; Parsons, 2015).

Chapter Summary

In this chapter we have seen how machines can learn from data and make complex predictions and decisions based on that knowledge acquired from data (McCarthy, 2007). Due to the ever-increasing amount of data in and the higher data processing capacity of modern computers, the interaction of data, machines, and humans will soon increase even more (Shaw et al., 2019; Veale, 2020). A substantial step toward large-scale processing of vast and complex data by a machine will be the self-driving vehicle, which is thought to transport its occupants autonomously even in unknown and ambiguous situations thanks to a multitude of sensors and complex algorithms (Bechtel et al., 2018; Bergmann et al., 2018).

Still, self-driving vehicles are facing technical and legal challenges (Greger, 2018; Hevelke & Nida-Rümelin, 2015b). It is not yet possible to use self-driving vehicles without a human supervisor (IT Commitee of the United Nations, 1968). On the one hand, this is prohibited by the legal framework since it is unclear who should be held accountable in the event of an accident (Greger, 2018). On the other hand, the development of the ADVs' underlying technology is not advanced enough to cope with the uncertainty and complexity offered by the real world (F. Chen et al., 2020; Ramos et al., 2017).

Nonetheless, self-driving vehicles are a desirable technology because they are in principle able to transform our society and cities toward more inclusive and sustainable transport (Chehri & Mouftah, 2019). The automation of the vehicle would also offer a new opportunity for equal rights for vulnerable road users, who have so far been left behind or even left out in road traffic (Cusumano, 2020)(Chehri & Mouftah, 2019). Despite all remaining issues, this technology is desirable and offers added value for society (Faulhaber et al., 2019).

Another obstacle for the realization of this technology is a lack of acceptance among potential customers (Howard & Dai, 2014). It is vital to increase knowledge about the opportunities and risks to create a high level of acceptance (Benleulmi & Blecker, 2017). This is due to the fact, that the acceptance and willingness to use ADVs will shape the extend to which ADVs will be used in individual transportation (Krueger et al., 2016). Only if a majority of society trusts ADVs to drive passengers safely, this technology will be used in a large scale. How to increase acceptance raises new problems, like how an ADV should communicate with other road users (Gillath et al., 2021)? How should an ADV communicate adequately with its passengers to enable trust building? Previous research could already show that trustworthy communication is vital in the case of automated, but not fully autonomous traffic, since the irony of automation can occur (Bainbridge, 1983; Schreurs & Steuwer, 2015): People who are supposed to monitor a system and intervene in unclear or life-threatening situations can either not do so because the task of driving became too complex or because the supervisor is so far out of the control loop that they are unable to react fast and adequately (Endsley & Kiris, 1995; Jarosch, Bellem, et al., 2019).

This is precisely where this thesis proposes a new form of human-machine interaction that can potentially resolve the named issues. Therefore, we will propose an adaptive HMI that is thought to decrease reaction times, while increasing precision in takeover situations with tailored information about current traffic situations. Additionally, the provided information is thought to increase trust in the automated system and therefore the willingness to use an ADV (Koo et al., 2015; Othman, 2021). With newly developed methods we want to examine how the unaware driver can get back into the control loop fast. The final goal of this thesis is to propose a design for a human-machine cooperation that includes reciprocal representations of user and machine (Altendorf et al., 2017). This not only improves interaction in terms of a safe usage, but also improves trust and acceptance as the system adapts to the user and provides explanations for the system's decisions.

The following chapter presents a virtual world in which it is possible to test automated systems under to support people in critical traffic situations under realistic, safe and controlled conditions.

2

LoopAR: Human-Machine interaction during take-over requests

Nezami, F. N.†, Wächter, M. A.†’, Maleki, N., Spaniol, P., Kühne, L. M., Haas, A., Pingel, J. M., Tiemann, L., Nienhaus, F., Keller, L., König, S. U, König, P. & Pipa, G. (2021). ***Westdrive X LoopAR: An Open-Access Virtual Reality Project in Unity for Evaluating User Interaction Methods during Takeover Requests.*** **Sensors, 21(5), 1879. doi: 10.3390/s21051879**

† shared first authorship
’ corresponding author

Abstract

With the further development of highly automated vehicles, drivers will engage in non-related tasks while being driven. Still, drivers have to take over control when requested by the car. Here, the question arises, how potentially distracted drivers get back into the control-loop quickly and safely when the car requests a takeover. To investigate effective human–machine interactions, a mobile, versatile, and cost-efficient setup is needed. Here, we describe a virtual reality toolkit for the Unity 3D game engine containing all the necessary code and assets to enable fast adaptations to various human–machine interaction experiments, including closely monitoring the subject. The presented project contains all the needed functionalities for realistic traffic behavior, cars, pedestrians, and a large, open-source, scriptable, and modular VR environment. It covers roughly 25 km^2, a package of 125 animated pedestrians, and numerous vehicles, including motorbikes, trucks, and cars. It also contains all the needed nature assets to make it both highly dynamic and realistic. The presented repository contains a C++ library made for LoopAR that enables force feedback for gaming steering wheels as a fully supported component. It also includes all necessary scripts for eye-tracking in the used devices. All the main functions are integrated into the graphical user interface of the Unity® editor or are available as prefab variants to ease the use of the embedded functionalities. This project's primary purpose is to serve as an open-access, cost-efficient toolkit that enables interested researchers to conduct realistic virtual reality research studies without costly and immobile simulators. To ensure the accessibility and usability of the mentioned toolkit, we performed a user experience report, also included in this paper.

2.1 | The Out-Of-The-Loop Unfamiliarity

Introduction

What defines the user-friendly design of automated systems has been the subject of scientific discussion for decades (Bengler et al., 2020; Norman, 1990). Especially in the upcoming years, when automated vehicles of SAE (society of automotive engineers) automation levels 3 and 4 will emerge, the demands on the driver's cognitive system will alter radically, as the role of humans as continuously active decision-makers in vehicles is replaced by automated systems (S. Li et al., 2019; Lindgren et al., 2020). Such techniques include the Audi traffic jam pilot (Audi AG, 2021) or Tesla's full self-driving beta (Tesla Motors, 2020). Airlines' experiences, where automated systems are already widely integrated, clearly state that such systems' safety and reliability cannot be achieved by optimizing technical components alone (Masalonis et al., 1999). Instead, the reliability of highly automated systems is primarily determined by the driver's cognitive processes, meaning how fast a safe transition to manual drive is possible (Zeeb et al., 2015) .

The need for a fast and safe transition applies particularly to situations where humans have the task of taking over system control in the event of sensor failures or malfunctions (Abe et al., 2011; Maurer, 2015). Thus, investigating the fluent integration of the takeover request (ToR) is crucial for the safety of any system with even partially automated driving features (Marberger et al., 2018). During a takeover request, the human driver most likely has to take over control in under 10 s, even when not engaged in driving-related activities (Dogan et al., 2019; Gold et al., 2013; Melcher et al., 2015). Naturally, an orientation phase follows as the human driver has to assess the traffic situation (Gold et al., 2013). Unfortunately, the driver's reaction is often too slow in critical situations, potentially resulting in an accident in the small time frame (<4 s) before an impact occurs (Green, 2000; Summala, 2000). Even in the case of fast reactions within a time frame under 10 s, studies with prolonged driving have shown hectic responses by human drivers, which of course neither improved the reaction time nor the situational outcome (Endsley & Kiris, 1995; Jarosch, Bellem, et al., 2019).

This manuscript presents a new toolset for human–machine interaction research apart from typical screen-based simulators. Existing simulators are often based on actual car interior designs. Therefore, they offer only limited possibilities for human–machine interaction (HMI) research (Morra et al., 2019). A very similar problem is posed by research on prototype cars in the real world, where realistic accident scenarios are costly and can only be generated to a minimal extent without endangering the test person involved. The project, called LoopAR, provides not only all the needed assets and an environment but also all the needed code to display the information of a takeover request as a freely programmable augmented reality (AR) feature in the windshield. The developed HMI displays the takeover request and highlights critical traffic objects to enable participants to take over more quickly and precisely. Our research is aimed toward safe and effective communication between car and driver. This is not only beneficial in terms of safety for the passengers but could also increase customer acceptance of highly automated vehicles, since up until now, malfunctions have been vital concerns of possible customers (Howard & Dai, 2014). Since LoopAR is based on the project Westdrive (Nezami et al., 2020), all the code needed and designed scenes are available in a Github repository. Project

Westdrive is an open science VR project that tries to enable many researchers to conduct VR studies. It provides all the necessary code and assets in a public repository to set up VR studies. LoopAR is an extension of the Westdrive toolkit, focusing on the human–machine interaction. To fully use the project presented here, only a powerful computer, VR glasses, a simulation steering wheel and pedals, as well as Unity as a development program are required.

2.2 | A VR Toolkit for the Human-Machine Interaction

Methods and Main Features of LoopAR

The main focus of the presented project is versatility and modularity, which allows the fast adjustment of the environmental and functional objects via prefab and the provided code in the toolkit. Research on the interactions between humans and cars is mostly done with stationary simulators. Here, a whole car chassis is used, or only the interior is set inside a multi-screen setup. However, these classical setups are often expensive, and adjustments or graphical improvements of the stimuli used in an experiment are often not possible (Cruden, 2018). In the past few years, there has been a significant shift in research toward virtual environments. This is reflected by applications like Cityengine and FUZOR (ArcGIS, 2013; Kalloc Studios, Inc, 2013) and by the software for driving environments (Dosovitskiy et al., 2017).

Still, experimental designs on human–machine interaction, in terms of specific car interior adjustments, are not possible yet. Therefore, the presented project enables the user to create experimental conditions and stimuli freely. All functionalities that are mentioned in the following are independent and can be adjusted at will. Additionally, the presented project does not need a specific hardware setup, making it easily adaptable and future-proof. New components, e.g., new GPUs and new VR devices, can be easily integrated into the setup displayed in Figure 2.1. The current requirements only apply to the VR devices used and are not bound to the toolkit. The following figure depicts an overview of the default experimental procedure, environmental structure, and data flow of the toolkit. Again, all of these defaults can be adjusted at will. The configurations presented here are intended to allow for a quick adaptation to other experiments.

Platform

Project LoopAR is made with the Unity editor 2019.3.0f3 (64bit). This software is a widely used game engine platform based on C# by Unity Technologies, supporting 2D, 3D, AR, and VR applications. The Unity editor and the Unity Hub run on Windows, Mac, and Linux (Ubuntu and CentOS), and built applications can be run on nearly all commercially usable platforms and devices. Unity also provides many available application programming interfaces and is compatible with numerous VR and AR

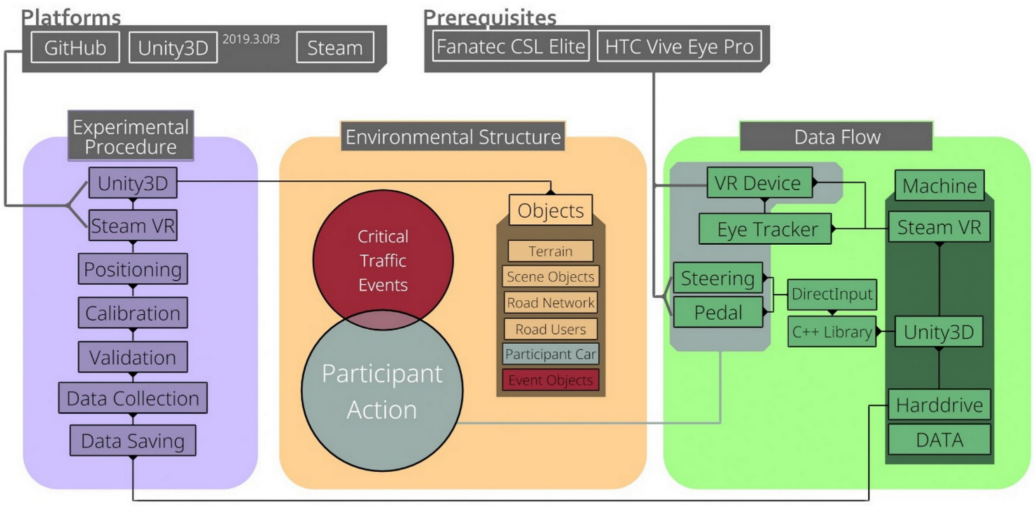


Figure 2.1: A simplified overview of the toolkit structure. It includes the default experimental procedure, a possible example of how the environmental structure can be used, and the standard data flow of the toolkit.

devices (Juliani et al., 2020).

The backend code of the project LoopAR was developed entirely using C# within Unity3D Monobehaviour scripting API. The backend comprises functionalities including dynamic loading of the environment, AI car controls, pedestrian controls, event controls, car windshields augmented reality controller, data serialization, and eye-tracking connection. Additionally, the presented project contains a C++ library enabling the force feedback for Microsoft DirectX devices that enables various force feedback steering wheels to function as controllers altogether. LoopAR code has been developed with modularity in mind to avoid complicated and convoluted code. All functionalities can be enabled or disabled individually using the Unity editor's graphical interface based on need.

Virtual Environment

To test human–machine interactions, an interactive and realistic 3D environment is needed. LoopAR aims at a fully immersive experience of a highly automated car encountering critical traffic events. To be able to investigate different driving conditions and scenarios, we created four independent scenes. In the following section, the environment design decisions are presented together with a short description of the experimental scenes.

The LoopAR environment is based on real geographical information of the city of Baulmes in the Swiss Alps. We selected this region due to its variety of terrain,

Figure 2.2: LoopAR map preview: mountain road (3.4 km), city (1.2 km), country road (2.4 km), and highway (3.6 km).

including a small village, a country road, a mountain pass, and a region suitable for adding a highway section, totaling around 25 km² of environment and an 11 km continuous drive through different roads. To reduce the computational demands, we sliced the terrain into four areas. Due to the road network design, these separate environments can be merged (Figure 2.2). These areas demand different driving skills from an automated driving vehicle and a human driver, reacting in different situations with different conditions according to the landscape and traffic rules. To make the region accessible in Unity, we used the collaborative project OpenStreetMap (OSM)(OpenStreetMap Foundation, 2004) and the open-source 3D software Blender (Blender Foundation, 1995).

OpenStreetMap is a project with the aim of creating a free map of the world. It collects the data of all commonly used terrains on maps. The project itself collects information, so the data are free of charge. The virtual environment contains a mountain road scene (Figure 2.3), including curvy roads winding through a forest and steep serpentines running down a mountain. These curvy roads require various driving speeds (from 30 km/h or slower, up to 100 km/h on straight stretches). The overall traffic density is low.

The second area of the environment is the village "Westbrück" (Figure 2.3). Here, it is possible to test events in a more inhabited environment. This environment is

(a) (b) (c)

Figure 2.3: (a) Pictures of the different scenes from the mountain road. (b) Pictures of the different scenes from the village "Westbrück". (c) Pictures of the different scenes from the country road.

characterized by narrow streets and dense traffic in low-speed environments.The third scenario is the country road scene (Figure 2.3), designed for medium to high speed (70 km/h), medium traffic density, and a long view distance. The last scenario for the participants is the highway scene (Figure 2.4), enabling critical traffic events with a higher speed and a low to medium traffic density.

Critical Traffic Events

To test the participant's behavior in critical traffic events, we created limited event zones, where the monitoring of a participant can be achieved in a well-controlled

(d)

Figure 2.4: (d) Pictures of the different scenes from the highway.

environment. In Figure 2.5, one example of a traffic event is displayed. Each environment (mountain road, city, country road, and autobahn) has three critical traffic events. These zones are the core of the possible measurements in the presented toolbox. Simply put, the event system is realized by a combination of several trigger components. These independent triggers are activated when the participant enters the start trigger (Figure 2.5: green gate). The event zone is restricted within “boundary” triggers (Figure 2.5 : yellow boxes). These triggers get activated on contact, which is considered a participant’s failure. Contact with the event triggers leads to a black screen followed by a respawn of the car at a point after the event (Figure 2.5: pink box) and giving back the car’s control. An event is labeled as “solved” when the participant enters the end trigger (Figure 2.5: red gate) without crashing, i.e., making contact with the “boundary” triggers. All critical events can be adjusted at will, and a prefabricated file is stored in the repo to create new events. The triggers are all visible in editor mode but invisible to the participant.

Cars and Traffic Behavior

Within the event zones, dynamic objects, such as other road users, are needed to create realistic traffic scenarios. The repository presented here contains various animated pedestrians, animals, and cars to create a broad range of critical situations. Additionally, there are some busses and trucks, and some construction site vehicles that can be used. Furthermore, a user’s own fbx models, as well as vehicles from the Unity asset store, can be added. For more details, please see the Supplementary Materials. All cars used are based on the Unity wheel collider systems of the Unity3D physics engine. In the Car Core Module, user input is translated into the motor control of the participant’s car. The input consists of the motor torque, brake torque, and steering, which are applied to the wheels. This functionality is called AI control. It allows a seamless transition from automated to manual driving when activated.

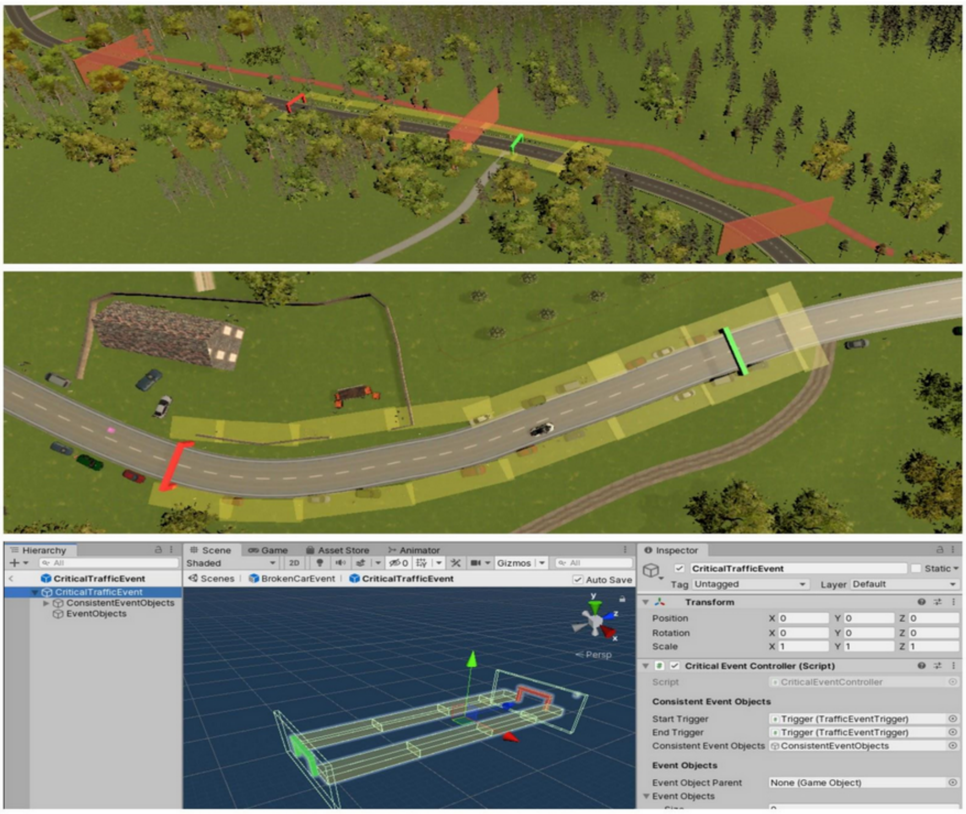


Figure 2.5: Traffic event prefab and its implementation.

To facilitate realistic traffic behavior, an additional AI module enables cars to follow predefined paths. Paths followed by AI Cars and walking pedestrians were defined by mathematical Bézier curve paths (Prautzsch et al., 2002), which were realized by the Path-creator tool (Lague, 2021). Speed limit triggers inside the scene manipulate the AI's aimed speed, handling the input propagated to the Car Core Module. Another module of the car AI allows the AI cars to keep a distance from each other. The goal is to create an easily configurable and interchangeable traffic AI for multiple study designs. With these measures, we maximized the car physics and traffic simulation realism while ensuring easy adjustments.

Experiment Management

Data sampling, dynamic objects, and driving functionalities within the event zones are controlled by a system of experiment managers that handle scene-relevant information and settings shortly before and during the real experiment phase. It handles different camera settings, the information given by triggers inside the scene, and the participants' respawn in case of failure. Before an experiment starts, initial adjustments start the experiment. These adjustments configure the experiment to the participant and include the eye calibration, eye validation, seat calibration, and a test scene. The eye-tracking component in this setup comprises an eye-tracking calibration, validation, and online gaze ray-casting, which can record necessary gaze data during the experiment. The component was built for the Tobii HTC Vive Pro Eye device but is intended to keep the VR component interchangeable. It was designed as a simple connector to tap into SRanipal and the Tobii XR software development kit (SDK) (Figure 2.6). The eye calibration is performed with the built-in Tobii eye calibration tool. The validation is set in the corresponding validation scene, which provides a simple scenario with a fixation cross. Validation fails if the validation error angles exceed an error angle of 1.5° or the head is moved by 2" from the fixation cross. During the experiment, the eye orientation, position, and collider hits are stored with a calculated gaze ray of both eyes. Currently, it is set to receive information about any object inside these rays to prevent the loss of viable information by objects covering each other.

In addition to the eye-tracking data, input data of the participant as well as scene-relevant information, such as the number of failed critical traffic events, are saved using generic data structures and Microsoft Linq, serialized into JavaScript object notation (JSON), and saved with a unique ID at the end of each scene. The generic data structure used in the project ensures flexibility, as different data types can be added or removed from the serialization component. This approach guarantees the highest compatibility with varying analysis platforms such as R or Python for the data gathered with LoopAR.

By conducting data saving, and given the nature of the experimental setup, we aim for a stable and high frame rate to provide a less sickness-inducing experience. A stable visual experience can be seen as a prerequisite to avoid potential sickness (LaViola, 2000). The desired optimum for the experiments is a stable frame rate matching the fixed rate of 90 Hz used by the manufacturers HTC and Oculus. Our current frame rate in the different scenes yields an average of 88 samples per second in our test setup, matching the maximum sampling rate of the HTC Vive with 90 fps.

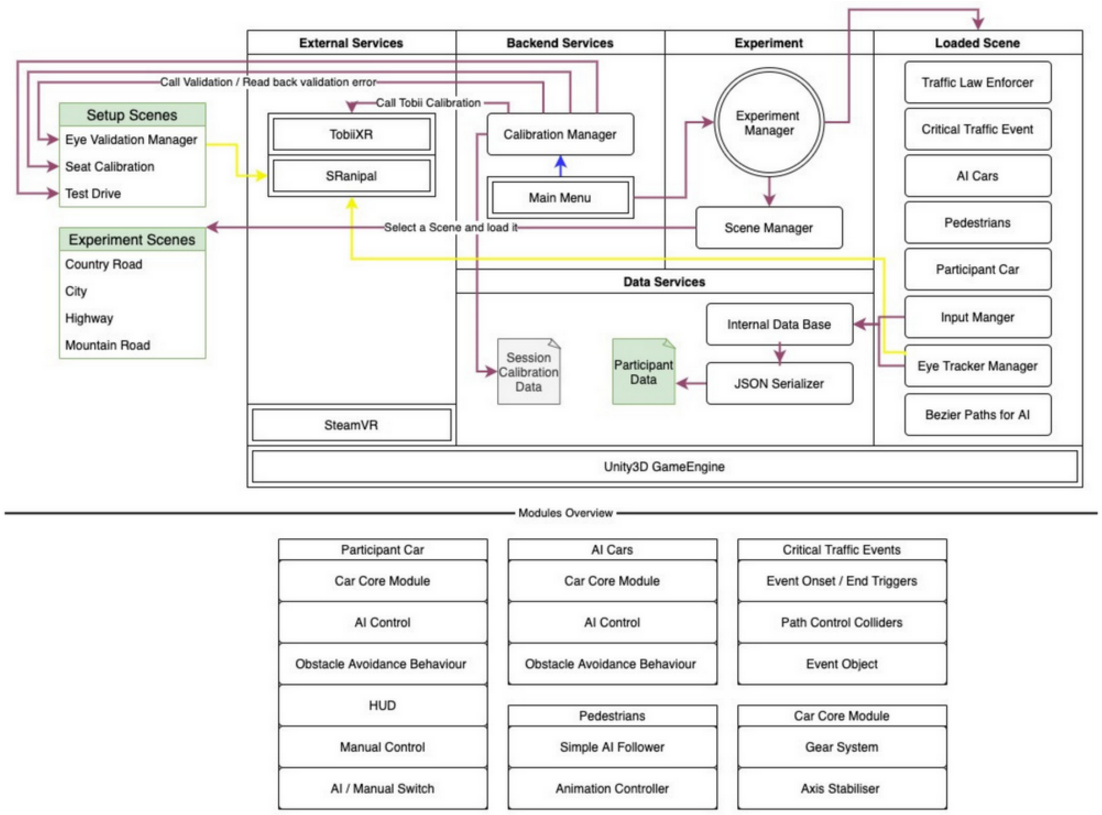


Figure 2.6: Scheme of the LoopAR functionalities and components illustrating the interaction of the different services and manager scripts within the Unity environment.

Hardware Requirements

The setup used and presented here is thought to be a cost-efficient and very mobile replacement for maintenance-intensive, rigid, and expensive driving simulators for studies on human behavior in the context of self-driving cars. A key advantage is freedom regarding the selected components. The only requirement for operation is granting the computing power for the entire system, which consists of a core setup only of a computer, a head-mounted display, and a steering wheel (see Table 1).

As a virtual reality device, we used the HTC Vive Pro Eye with an integrated Tobii Eye Tracker. It is a cable-bound head-mounted display that enables the participant to transfer movements into virtual reality. Although we are using the Vive Pro exclusively at our department, the LoopAR experiment is not dependent on this specific VR device. We used the components of the setup with 90 fps sampling and display.

2.3 | Westdrive X LoopAR Usability

Discussion

In the presented paper, we describe LoopAR as a modular toolkit to test a takeover of control in critical traffic situations from automated cars to human drivers by combining VR and eye-tracking in an interactive and immersive scenario. Its current state and design provide a promising, new, low-cost, and mobile setup to conduct studies that were traditionally only done in stationary simulators. The current code, as well as the 3D environments, can be adjusted at will. With newly implemented code, it is not only possible to simulate a large and highly realistic VR environment, but it is also possible to create a broad range of applications in VR research that is not only bound to HMI investigations. A large part of the assets used are from Unity's asset store and the 3D platforms Sketchfab and Turbosquid. Therefore, it is possible to change the number, size, and shape of all objects in each scene.

All of the functionalities above, and assets presented here, are under constant improvement. By writing, five new projects, ranging from ethical decision-making over EEG implementation to human spatial navigation, arise from the presented toolkit, which will also develop new assets and features implemented into the toolkit later on. The authors want to emphasize the modularity and adaptability of this VR toolkit.

User Reports

To check for the user friendliness of the presented toolkit, a system usability score (SUS) -based report was performed (Lewis, 2018). Here, we asked 11 of the current users between the age of 23 and 34 (5 female) to evaluate the usage of the main features in the toolbox starting from cloning the repository, adjusting the environment, and manipulating dynamic objects in an example scene.

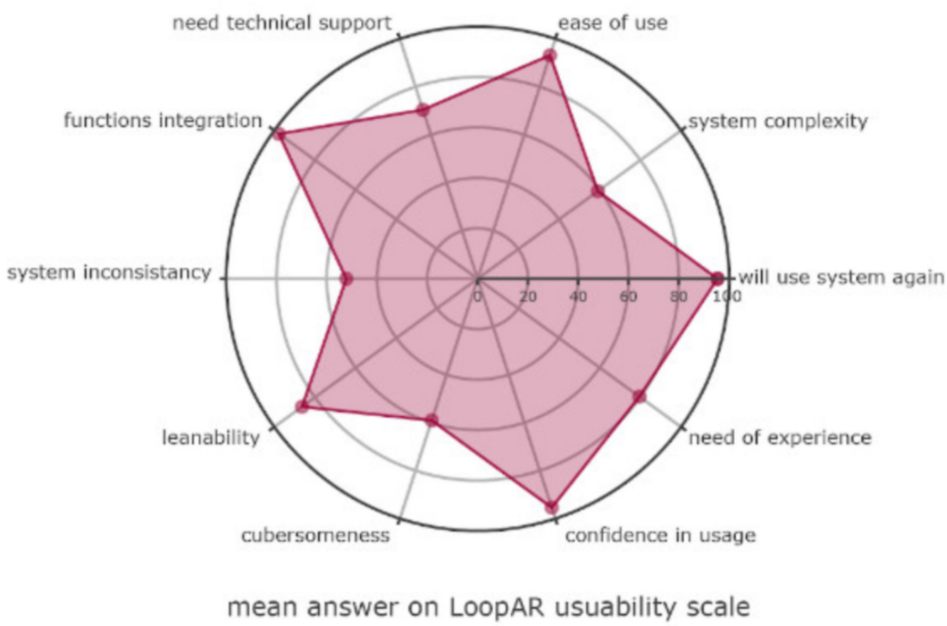


Figure 2.7: Visualization of the usability report items as a radar plot of the system usability scale data

While doing so, we asked the participants to evaluate the feasibility of the tasks. User experience in Unity and C# programming varied from no experience to expert levels with more than 3 years of experience. Our top findings, depicted in indicate that the toolbox is perceived as well documented, and advanced Unity users faced no major problems building and altering their project created with this toolbox (Figure 2.7, Figure 2.8, Figure 2.9). While some steps in the procedures might be challenging to new users, the Westdrive X LoopAR toolbox seems to be a useful foundation for all users.

Figure 2.8: Visualization of the usability report items as a word cloud showing most frequently used words in the comments

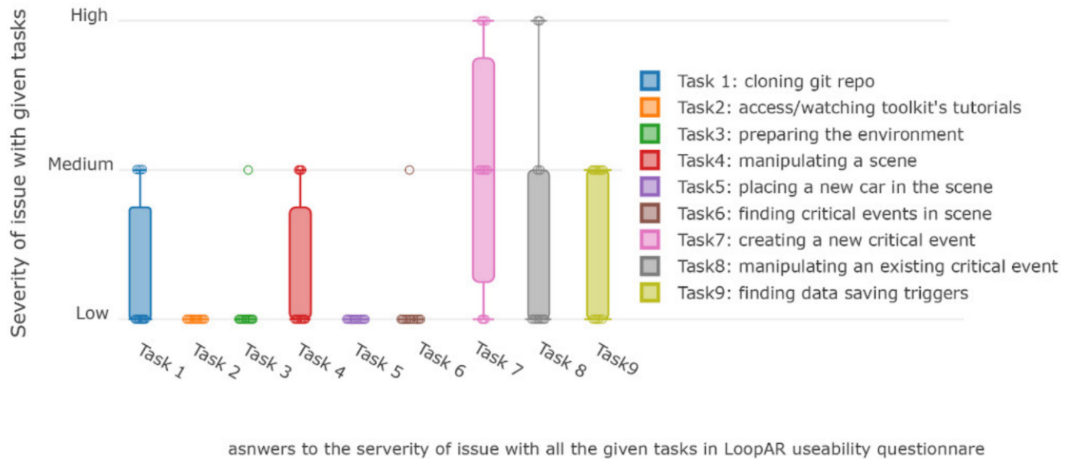


Figure 2.9: Visualization of the usability report items with a severity of issue bar plot, related to the tasks in the usability report. Low equals no delay in time or perceived obstacles, medium refers to a completed task with added effort. High indicates noticeable delay or frustration and that the participant may not be able to complete the task.

Conclusion

This article describes a new virtual reality toolkit for Unity applications investigating human–machine interaction in highly automated driving developed by us. The presented setup is thought to be a mobile, cost-efficient, and highly adaptable alternative to chassis simulators that closely monitor the participants. It is particularly noteworthy that there is not only a drastic reduction in costs but also an improvement to the adaptability of the software as well as the used hardware. All components are fully upgradable, in case there are better products in terms of image quality or computing power. LoopAR allows interested researchers to conduct various virtual reality experiments without creating the needed environment or functionalities themselves. For this, we have provided an area of almost 25 km^2 based on OSM data. The toolkit presented here also includes all the necessary assets and basic prefabs to quickly and precisely create a wide variety of virtual environments. Additionally, the LoopAR toolkit contains components of the experimental procedure and data storage.

Supplementary Materials

The following are available online at https://www.mdpi.com/1424-8220/21/5/1879/s1, Unity® 3D: www.Unity3d.com; Online Character animation: www.mixamo.com; Adobe Fuse CC: www.adobe.com/products/fuse.html; Blender 2.81: www.blender.org.

Author Contributions

F.N.N. and M.A.W. wrote this paper. Both authors designed the project. S.U.K., P.K., and G.P. supervised the LoopAR project. N.M. developed major parts of the AI, functional modules, and User Interfaces. J.M.P. realized scene building and the head-up-display (HUD) functionalities. L.K. designed the mountain road scene and provided performant assets. L.T. developed large parts of the software architecture and acted as a software engineer for the project's functional compartments. A.H. designed the highway scene and provided additional assets. L.M.K. designed the country road scene, provided assets, and contributed to HUD-related literature background. P.S. was involved in designing the city scene, as well as managing and creating assets. F.N. developed and designed the test drive scene. Additionally, we would like to thank Debora Nolte, Shadi Derakhshan, and Vincent Schmidt for their valuable user feedback. All authors have read and agreed to the published version of the manuscript.

Chapter Summary

This chapter introduced LoopAR as a virtual environment for testing multimodal human-machine interaction. It is thought to support participants in critical traffic situations under realistic, safe and controlled conditions by providing warning sounds and guiding the users's attention toward relevant objects in the surrounding environment. It was shown that realism, size, modularity and conformity with traffic regulations are factors that should be considered when designing such a virtual environment. Since LoopAR is a Unity toolbox, emphasis was on easy handling and open access under Creative Commons.

As seen in the first chapter of this thesis, increasing automation poses questions about how the unaware driver can get back into the control loop fast in case the car demands a takeover in a critical situation (Bainbridge, 1983; Endsley & Kiris, 1995). Existing studies could show that especially during fast takeover maneuvers, where control has to be handed form the vehicle to the human user within seconds, the initial warning displayed to the human should be presented very clearly and unequivocally across multiple modalities (Jarosch, Bellem, et al., 2019; Merat et al., 2014; Norman, 1990; Roche et al., 2019) to prevent any potential time loss in the transition. This is necessary, since the driver can be engaged in non-driving related tasks and therefore could be distracted by music, reading or even both.

A warning in a single modality would not be sufficient to initiate a successful takeover request (Jarosch, Bellem, et al., 2019; Marberger et al., 2018). The request also has to be precise, since the transition is often time critical transition from a non-driving related task back to driving takes longer when the non-driving related task is one that includes a motor component, such as holding a cell phone or a tablet (Jarosch, Bellem, et al., 2019). Research in takeover scenarios are mostly done with simulator setups. However, simulator setups often imply high technical and financial expenditures (Zhang et al., 2019). Additionally, simulator setups are usually based on actual existing vehicles and thus, do not allow an independent design of the HMI. We therefore wanted to show that proper research can be achieved in a versatile and cost-efficient simulator replacement using customer-grade VR equipment. Additionally, we presented the code needed and the assets used in this environment that enable interested users and researchers to easily and quickly adapt the toolkit to their needs, including the possibility of close monitoring of the subject via the integrated eye-tracker in the HMD. Therefore, LoopAR is a toolbox that allows research on a wide range of human-machine interaction scenarios, investigating the OOTLU and possible counter measures.

After the publication of LoopAR as toolbox, the next step of the project, was data acquisition. During autumn and winter 2020, we measured more than 200 subjects from a broad distribution of society in 2020 in order to conclude whether visual and auditory stimuli enhance precision of a human action during takeover. The data of our subjects are stored within the university. A cursory review showed that data seems to be are consistent with the existing literature. We suggest, that the multimodal HMI decreases reaction times while increasing the probability of navigating successfully through the critical traffic situation. The ultimate analysis will be much more profound since the recorded data will look at the outcome of the different events, the angular position of the pedals and the steering wheel, and the head and eye movements. The plan of the pending data analysis is to gain insight into which factors play a role in the takeover request and which experimental conditions can lead to a faster and more precise takeover situation of level 3 advanced driver assistance systems.

Level 3 automation is currently a point of interest in research and development, since ADVs of higher levels 4 and 5 are facing substantial obstacles (Greger, 2018). The legislation on the safe vehicle operation stipulates that electronic, traffic-relevant operations must be redundant and injury-free (Bundesministerium für Digitales und Verkehr, 2017). These requirements have massive implications for the construction of future vehicles. For example, in automatic mode, the steering wheel must be decoupled from the steering angle of the wheels to prevent injury to the driver if the vehicle makes a rapid steering movement. However, if the steering wheel is decoupled, the driver cannot directly start the vehicle takeover. Large-scale implementation of this technology seems complicated under these circumstances, as the human-machine interaction is faced with many similar unsolved issues (Waschl et al., 2019). At this point it should already be considered whether the partial automation might be a suboptimal step towards full automation. It is conceivable that models that pose fewer legal issues, such as level 2 or 3 as partial automation, will become a standard in future automobiles. Therefore, a new form of cooperation between a human and highly automated vehicle is needed to ensure safe operation of partially automated vehicles.

But even if it is not yet clear whether self-driving vehicles of levels 4 will be approved for public use at all, it makes sense to start thinking about how such technology can be designed to guarantee a proper interaction between humans and machines. Therefore virtual designs and tests can not only be used to avoid potential weaknesses before the technology is realized, but also possibly improve the user experience in levels 2 and 3 (Waschl et al., 2019).

In level 4, the vehicle needs to be able to transfer itself to a safe state and is not dependent on the human as a supervisor, who must be ready for a takeover in the event of a hazard reaction (SAE Internation, 2014). However, this also means that the driver has less control over the vehicle in general. This raises the question of the extent to which potential customers trust such technology since they no longer directly influence lateral and longitudinal acceleration.

Since the driver is not able to take direct control in automation level 4, the relationship between the human and machine changes substantially (Bainbridge, 1983; Bengler et al., 2020). The driver becomes more of a passenger, since he has to trust that the vehicle is acting in best interest during the trips in which the automation features are engaged (SAE Internation, 2014). Therefore, it is vital to increase acceptance through features in the interior.

The following chapter will examine human acceptance and trust of self-driving vehicles in the context of level 4 and 5 vehicles. More specifically, we will concentrate on verbal feedforward and feedback to see if the communication strategy does have an effect on the acceptance of self-driving cars, assessed via a post-experimental questionnaire as well a behavioral data of the head movements. The experimental setup that will be discussed also includes a VR toolbox in the context of a large city. Before we discuss possible communication strategies of self-driving vehicles in chapter 4, we will first discuss the virtual environment used for this purpose in the next chapter.

3

Motor City: A virtual reality foundation for Human-Machine research

Nezami, F. N.†, Wächter, M. A.†', Pipa, G., & König, P. (2020). ***Project Westdrive: Unity city with self-driving cars and pedestrians for virtual reality studies.*** **Frontiers in ICT, 7, 1. doi: 10.3389/fict.2020.00001**

† shared first authorship
' corresponding author

Abstract

Virtual environments will deeply alter the way we conduct scientific studies on human behavior. Possible applications range from spatial navigation over addressing moral dilemmas in a more natural manner to therapeutic applications for affective disorders. The decisive factor for this broad range of applications is that virtual reality (VR) is able to combine a well-controlled experimental environment together with the ecological validity of the immersion of test subjects. Until now, however, programming such an environment in Unity® requires profound knowledge of C# programming, 3D design, and computer graphics. In order to give interested research groups access to a realistic VR environment which can easily adapt to the varying needs of experiments, we developed a large, open source, scriptable, and modular VR city. It covers an urban area of nearly 2,5km^2, up to 150 self-driving vehicles and 655 active and passive pedestrians and thousands of nature assets to make it both highly dynamic and realistic. Furthermore, the repository presented here contains a stand-alone City AI toolkit for creating avatars and customizing cars. Finally, the package contains code to easily set up VR studies. All main functions are integrated into the graphical user interface of the Unity® Editor to ease the use of the embedded functionalities. In summary, the project named Westdrive is developed to enable research groups to access a state-of-the-art VR environment that is easily adapted to specific needs and allows focus on the respective research question.

3.1 | A Toolkit for Virtual Reality Experiments

Introduction

With the opening of the consumer market in recent years, VR has penetrated many areas of everyday life: there are e.g., applications for marketing, the games industry and for educational purposes (Anthes et al., 2016; Burke, 2018; A. Miller, 2018). Research on human behavior is also beginning to take an interest in experiments in virtual reality (de la Rosa & Breidt, 2018; Rus-Calafell et al., 2018; Wienrich et al., 2018). For instance, it is possible to embed ethical decision making in a seemingly real context in order to achieve a higher validity of experiments (Faulhaber et al., 2019; Sütfeld et al., 2017). Further, studies based on VR techniques address questions regarding spatial navigation, such as neurological correlations of human navigation (Epstein et al., 2017), as well as gender differences in navigation tasks in a well-controlled environment (Castelli et al., 2008). Although there are already available tools for creating virtual cities, these applications have not yet been designed for experiments on human behavior, but rather for planning and simulating urban development (ArcGIS, 2013; Botica et al., 2015; Kalloc Studios, Inc, 2013). Furthermore, it is possible to use VR in a variety of psychotherapeutic and clinical scenarios (A. Li et al., 2011; Riva, 2005). Not only is this cost-efficient and more interactive than classical psychotherapy (Bashiri et al., 2017), it also offers the possibility to use this treatment at home, as VR becomes more widespread in the future. This means that VR has the potential to increase access to insights of human behavior as well as to psychological interventions (Freeman et al., 2018; Slater & Wilbur, 1997). Finally, VR can be combined with further technologies, such as EEG (Bischof & Boulanger, 2003)and fMRI, facilitating research of clinical disorders (Reggente et al., 2018). In summary, VR techniques have the potential to heavily advance research in the human sciences.

Still, compared to classical screen experiments, VR-based experiments are complex and require extensive programming, which is an intricate task by itself (Freeman et al., 2018). This causes VR experiments in behavioral research to lag behind their actual potential (Faisal, 2017). Even if already existing experiments are transferred to VR, knowledge of software and hardware must be acquired, meaning a larger expenditure of time and content (X. Pan & Hamilton, 2018). Westdrive is developed to eliminate these obstacles in the context of studies on spatial navigation and ethical aspects. It shortens the time required for the setup of or the transfer to VR experiments by a considerable magnitude either by enabling researchers to use the project scene directly, or indirectly by letting them use only the provided assets and code.

3.2 | Relevant Features for Virtual Environments

Results

Key Features

Probably the most crucial features of Westdrive are size, modularity and the simple handling of complex environments, since all components of the City AI toolkit can be used independently even without any programming knowledge.

Size is often a critical factor for virtual environments. This is the case with e.g., navigation tasks within VR (S. U. König et al., 2019). A distinction is made here between room-sized vista space and large environmental space. Small rooms are easier to grasp and therefore it is only possible in large environments to distinguish between test subjects who navigate using snapshots of landmarks only and those who have learned a true map of their environment (Ekstrom & Isham, 2017).

The modularity of a virtual environment is of equal importance. Not only does building realistic cities require the consideration of many different aspects, but different research projects also depend upon distinctive dynamic objects. For example, an experiment on the trolley dilemma requires driving vehicles and pedestrians (Faulhaber et al., 2019). A therapeutic application for fear of heights requires high buildings and animated characters to make the environment appear real (Freeman et al., 2018). Project Westdrive offers a wide variety of applications due to its modularity, both of the static environment which comprises trees, pavements, buildings, etc. and the dynamic objects like pedestrians or self-driving cars.

Additionally, the aforementioned managers of the City AI toolkit enable a simple handling of the project. The City AI toolkit, which facilitates implementation of paths, pedestrians and cars, which are all usable within the Unity® GUI without any experience in coding. All of the components are accessible within the Unity® Editor. All managers can be edited separately according to the respective requirements of an experiment. In this sense, these separately adjustable components also support modularity as only adjustments for the necessary components have to be made.

To use the project, only a powerful computer, VR glasses and the free Unity® program are needed.[1] If the aforementioned requirements are met, the scene pre-

[1]GPU: NVidia GeForce GTX 1080 ti, equivalent or better, CPU: Intel(R) Xenon ® E5-1607 v4, equivalent or better, RAM: 32 GB, Video Output: HDMI 1.4, DisplayPort 1.2 or newer, USB Port: 1x USB 2.0 or better port, Operating System: Windows 8.1 or later, Windows 10.

sented here can be changed or manipulated at will. It is also possible to make alterations exclusively in the GUI of the Unity® editor without writing any code. This project offers not only the templates for static models, but also the functions integrated into the GUI for paths, character creation, and the creation of moving cars.

Westdrive and the City AI have been created with having simplicity in mind to relieve users from as much time-consuming preparations and programming as possible. Yet, as an open source project under constant development, we also encourage future researchers to further improve the project or change the codes based on their specific needs. Westdrive gives the user the possibility to carry out a multitude of investigations on human behavior through the key features. For example, the simple routing of pedestrians and cars makes it possible to carry out studies on trolley dilemmas or the human-machine interaction. Also, due to the realism of the avatars (Figure 3.1) it is possible to build therapeutic applications for the treatment of fear of heights or social phobias. However, this is only a very small part of the possible applications.

Figure 3.1: Overview of all used Fuse CC Avatars in the virtual city

3.3 | Technical Aspects of Motor City

Methods

Project Structure

The Westdrive virtual environment is built in Unity® 2018.3.0f2 (64 bit), a game engine platform by Unity® Technologies. This engine is used together with a graphical user interface (GUI) called the Unity® editor, which supports 2D and 3D graphics as well as scripting in JavaScript and C# to create dynamic objects inside a simulation. Unity® runs on Windows and Mac and a Unity®-built project can be run on almost all common platforms including mobile devices like tablets or smartphones. We have chosen this software due to many available application programming interfaces (APIs) and good compatibility with a variety of VR headsets (Juliani et al., 2020). Moreover, the use of Unity® grants access to an asset store, which offers the option to purchase prefabricated 3D objects or scripts which only need to be imported into an already existing scene. Thus, Westdrive is a modular virtual environment, making it easy to integrate other software now and in the future.

The Westdrive repository contains a city as one completed game scene. All associated assets including driving cars, walking characters, buildings, trees, plants, and a multitude of smaller 3D objects such as lanterns, traffic lights, benches etc. are included and offer a high level of detail (Figure 3.2). It also contains the relevant code that executes interactions and animations of the mentioned objects. Thus, users have all desirable components for an experiment in one consistent package. Westdrive can be divided into two sub-areas. On the one hand there is the static environment and on the other hand there is the code for interactions between dynamic objects. Both will be explained in the following.

Static Environment

The static environment models a large urban area. It includes 93 houses, several kilometers of roads and footpaths, about 10,000 small objects and about 16,000 trees, bushes and plants. A large part of the 3D objects used for this purpose are taken from the Unity® asset store for free. A list of used assets and their licenses can be found in the specified repository. However, the design of the city presented here can be varied at will in the editor and an included mesh separating tool. It is possible to change the size, shape and amount of individual buildings, streets, cars, and pedes-

Figure 3.2: Overview of the level of detail in the simulated city of the project Westdrive in a completed scene.

trians in the graphical user interface (GUI) of the Unity® Editor. The same applies to all other assets presented here. The static environment alone can thus be used indirectly for the development of further VR simulations as the project provides a large number of prefabricated assets (prefabs) that do not have to be created again. Consequently, it is possible to easily develop a broad range of scenarios for realistic VR experiments by simply manipulating the static environment to match respective needs.

Scripting Dynamic Objects

To use Westdrive to its full extent, the code described here is of essential importance. The code is written entirely in C# based on Microsoft's Net 4.0 API level and envelopes all functions for the stand-alone City AI toolkit (Figure 3.5). This includes six components developed by us: a Path Manager to create and manipulate paths for pedestrians and cars, a Car Engine script that allows cars to move independently, and a Car Profile Manager to create different profiles for different cars (e.g., the distance maintained to other vehicles, engine sound and car color). Additionally, there is the Pedestrian Manager and the Character Manager, that control animations and spawn points for moving characters along their defined route and an Experiment Profile Manager, which defines the experimental context, like routes, audio files, and scripted events along the path. The City AI works as a stand-alone toolkit in the GUI of the Unity editor. In short, it is possible to define fixed routes with spawn points for pedestrians and cars along which the non-player characters (NPCs)), such as pedestrians and cars, will move. Only if visual change of characters is desired an external tool is necessary.[2]

Figure 3.3: Impressions of cars in the highly realistic city scene.

To enable well-controlled movements of cars and pedestrians, we developed a path creation toolkit inside the City AI which incorporates mathematical components of Bezier Splines (Prautzsch et al., 2002). This results in a deterministic and accurate path following system which is only dependent on units of time in a non-physics-based simulation. The users can themselves change the control points of the path inside the editor (see Figure 3.4). It is also possible to define the duration of the route or the circuit in the Unity® editor. Furthermore, the kinematic path creation facilitates the creation of spawn points for different asset types (cars and pedestrians at the moment, see Figure 3.3, Figure3.1) for each path. All these functions work without programming knowledge. The following components of the City AI are also depicted in Figure 3.4 to give a better overview of interactions and possibilities within Westdrive:

[2]Character Creation: To create new characters for the city it needs two external tools. One is Fuse CC from Adobe and the other is the Mixamo website. In Fuse CC, a free 3D design program by adobe, it is possible to create figures according to your own imagination. This created model can then be uploaded to the program at Mixamo, which can automatically create animations (Aguiar et al., 2014). From this website, finished models with animations can be downloaded, which then only need to be implemented into the project. All characters created with Fuse CC in combination with animations from Mixamo can be used without licensing or royalty fees.

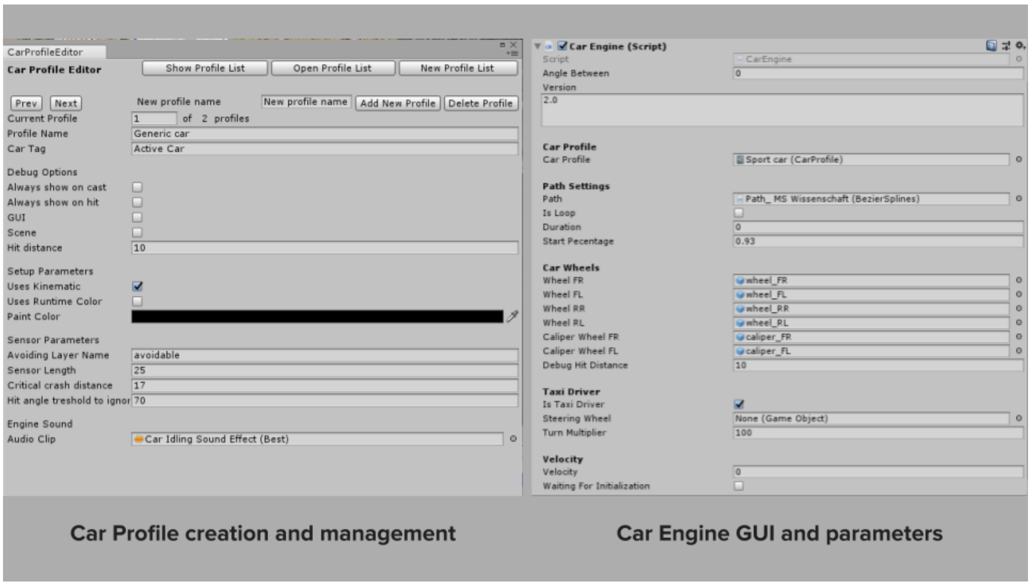


Figure 3.4: Overview in the Editor of the Car Profile Manager, the Car Engine and the according parameter bar. These functions allow users to use different types of vehicles in the city. The Car Profile changes the appearance of the vehicles, such as color, engine noise and sensor length. Car Engine allows the vehicles to move independently on the defined routes through the city and to accelerate, brake and steer independently. For each of these functions defaults are provided. An adjustment of these parameters is therefore only necessary for new vehicles.

Path Manager: This is the basis for all moving objects in Westdrive. With just a few clicks in the editor, the user can create new routes for pedestrians and cars or change existing routes. To do so, the control points of the already mentioned Splines can be moved using the mouse only. Afterwards it is possible to set the speed for objects on this route.

Car Engine: This component enables vehicles to steer, brake and accelerate independently both at traffic lights and in the event of an imminent collision with other road users.

Car Profile Manager: This component allows users to create and manage multiple independent profiles for cars. It enables creation of various types of cars with different characteristics such as engine sound, color, or a different spacer for preceding vehicles.

Pedestrian and Car Manager System: These systems take care of automatic spawn, restart, and re-spawn of all pedestrians and cars in the scene. They have the ability to load resources in an either synchronous or asynchronous manner, to ensure a smooth-running experiment.

Experiment Profiles and Procedure Controller: These scripts enable users to create different experiments within the environment. These profiles set up parameters for e.g., the routes that cars will follow. They also trigger the beginning and the end of the experiment; the end of the experiment blocks and they disable dynamic objects not necessary in the scene if needed. The Procedure Controller uses the Experiment Profile to automatize the experimental procedure e.g., by ending blocks, altering the appearance of or completely excluding dynamic objects.

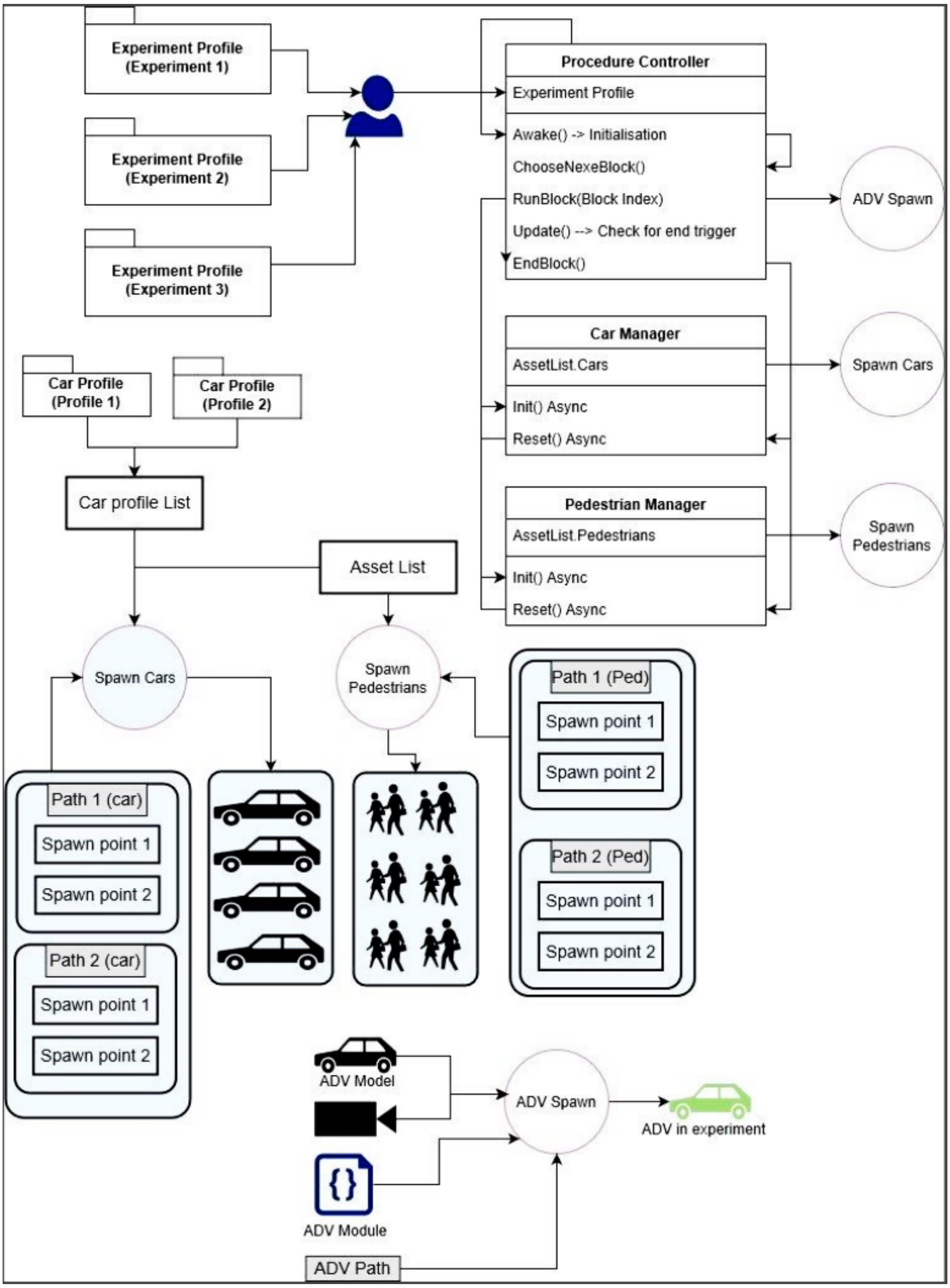


Figure 3.5: Scheme of the City AI features in Westdrive. This illustrates the interaction of the different managers of the toolkit to enable spawned cars and pedestrians as well as different experimental setups saved in one scene. These experimental profiles trigger the procedure controller, which takes care of the onset and ending of the experiment and creates the subject's car or avatar. This also triggers the car and pedestrian manager, which are responsible for the spawning of passive cars and pedestrians. In combination with the Car Profiles and the Asset List, the various cars and pedestrians required for the experiment are created in the experiment.

All of these managers assign the correct scripts to objects and move them to a resources folder in order for them to be spawned in runtime when the experiment starts. These toolkits ensure that cars and pedestrians have all the necessary components attached to them.

Implementation

As head-mounted display (HMD), the HTC Vive Pro is used at our department. At the time of writing, this virtual reality device is the most advanced technology available (Ogdon, 2019). In order to transfer the player's head movements into the virtual reality, HTC utilizes two passive laser-emitting "lighthouses" that have to be attached to the ceiling in two opposing corners of the room. The two handheld controllers and the headset use no <70 combined sensors to calibrate the positions of controllers and headset, measuring the time difference in sending and receiving the emitted signal (Ahmad, 2020). To use the HTC Vive Pro and the HTC Setup Software, an account at the online gaming platform Steam is necessary. This requires a stable internet connection, as both Steam and the HTC Setup software are free to use. Since this device is one of the most expensive ones on the market, it is used mainly for academic or industrial research rather than private gaming.

It is also worth mentioning, that although Westdrive has been developed for the HTC Vive Pro, it can easily be transferred to other virtual reality HMDs. The last component for the implementation is the Unity® software. Unity® can also be used free of charge as long as a project is not used commercially. Licenses are free for students and researchers. The Unity® editor can be downloaded from the Unity® website. Now it is possible to create a project order and convert the files from the repository presented here into Unity®.

A more detailed description of how to set up Westdrive as well as an example of the functionalities can be found as tutorial videos in the repository and in the Supplementary Materials.

3.4 | Limitations and Outlook

Current Limitations

Due to the complexity of the project and the differences between a deterministic simulation and a computer game, there are still many possible improvements to be implemented. With current enhancements like occlusion culling where, objects are not rendered when they are not seen by the player, simplified shadows, and mesh combining, an acceptable frame rate of at least 30 fps can be achieved using an NVidia GeForce RTX 2080 ti in combination with an Intel(R) Xenon® E5-1607 v4. The desired goal in the course of further research will be to reach the stable 90 Hz suggested by virtual reality technology providers such as HTC and Oculus.

It is important to note that the code does not calculate the mentioned objects physically, but kinematically, so no physically simulated forces are applied to any moving objects. There are several reasons for this: on the one hand, the computational requirements of the computer on which Westdrive is used on are kept as low as possible. On the other hand, an exact control bar of the visual stimuli can be guaranteed, because each object is spatially located exactly at the same place at the same time. Furthermore, it makes potential directed changes easy, as no physical interactions have to be reverse engineered. Another point is that there is currently no structured software architecture. So far, the priority has been on the simple handling of all functionalities within the editor to facilitate the creation of own experiments. A structured architecture is still under development.

Outlook

Concluding, we again want to emphasize the impact Westdrive can have on future VR research. Already over a decade ago, the potential of combining VR with physiological measurements has been discussed (Bischof & Boulanger, 2003), but only in the past years, when software became affordable, there was a renewed interest in VR in science (Interrante et al., 2018). The main advantage of the project is a simple implementation of a versatile project which, despite its complexity, can be altered quickly and easily without programming knowledge. Likewise, the experiment in its basic form doubles as an eye-tracking study. The code for the implementation is not included in this version, mainly because it was not written by the two authors, but by the Seahaven research group, investigating spatial navigation in a virtual environment (S. U. König et al., 2019). However, the repository will be constantly updated, thus it will also contain the required eye tracking code for Pupil Labs in the future. Westdrive as a city environment offers many areas of application. Nevertheless, the project is constantly in development and extension. At least two more scenes are currently planned in order to allow for an even wider application, for example the investigation of trolley dilemmas (Thomson, 1985)using a railway track or possible applications of the acceptance of new mobility concepts. All improvements and added scenes will be released via GitLab. Additionally, we are going to further clear up old parts of code and unused assets as code janitor, as well as fixing any possible typo or mistake in the code. At the same time, we will expand the comments and wiki section to have a user guide on how to use the project.

Since we are constantly improving the code and add functionalities, this cleanup is an ongoing process. In this work, particular importance was attributed to a comprehensible formulation in order to ensure an understandable documentation of the work performed. There is an almost unlimited number of application possibilities for the extension of this project. The authors are looking forward to the many great ideas for the continuation of Westdrive.

Author Contribution

MW and FN wrote this paper. Initial Idea to Westdrive began as a joint Master thesis. Both authors were building and designing the project. PK and GP supervised the project.

Funding

We acknowledge support by Deutsche Forschungsgemeinschaft (DFG) and Open Access Publishing Fund of Osnabrück University. This contribution is part of the research training group “va-eva: Vertrauen und Akzeptanz in erweiterten und virtuellen Arbeitswelten” of the University of Osnabrück.

Supplementary Material

The Supplementary Material for this article can be found online at:
https://www.frontiersin.org/articles/10.3389/fict.2020.00001/full#supplementary-material
Unity® 3D learning: www.unity.com/learn.
Online Character animation: www.mixamo.com.

Chapter Summary

As seen in the previous chapter, virtual environments offer a safe way to conduct scientific studies on human behavior in the context of self-driving vehicles. Importantly, virtual reality can combine a well-controlled experimental environment with a higher ecological validity than classical screen experiments due to the immersion of the subjects. Thus it is possible to create simulations of both level 3 and level 4 automation scenarios. Unlike the LoopAR project in Chapter 2, dealing with the transition of the DDT in level 3 automation, the current chapter is set in the context of level 4 automated vehicles, where the human passenger cannot intervene directly Shadrin and Ivanova, 2019. The vehicle still has a steering wheel, indicating that it is still possible to take control but the driver is not in the driver's seat, which prevents spontaneous intervention. Therefore, the presented toolkit focuses on the communication between car and passenger. The content of this toolkit comprises a large, scripted, and modular VR city build in Unity 3D. It covers an area of roughly 2,5 km^2, up to 150 self-driving vehicles, and 655 active and passive pedestrians, as well as thousands of environmental assets, to make it both highly dynamic and realistic. It is also possible to easily customize all avatars and vehicles even without dedicated programming experience. Although the creation of such a toolkit was not the first priority of the project, it was a useful side activity from our point of view. The reason for this is that although we saw the need for such a toolbox, as of 2019 there was still no free access to similarly complex toolboxes. In short, the projects described in this thesis and the resulting toolboxes should enable a broad range of interested researchers to test human-machine interaction in virtual environments.

Returning to the main objective of the thesis, the question arises how a virtual reality toolkit can be used to improve human-machine interaction in the context of self-driving vehicles, when humans are no longer actively involved in the driving process. In the depicted automation scenario, the occupant can no longer take immediate control. The distinctive difference between the LoopAR environment and MotorCity is that in the latter, there is no need for driving task support of the driver. As mentioned, this is due to the fact that the driver in level 4 automation is no longer part of the active DDT. Therefore, we disclosed the possibility of intervention for our study design and placed the participant onto the passenger's seat. As a passive occupant, the human can neither interfere with nor supervise the system. In these driving scenarios, trust in the decisions made by the vehicle plays a vital role. The reason for this is that the passenger must rely solely on the decisions of the vehicle. In order to build trust, the human being cannot gradually relinquish parts of the driving task in order to hand over control in small steps. Instead, the driver who has become a passenger must trust that the machine will act in his or her best interest.

So how exactly should a vehicle be able to generate trust if it can't make decisions overtly in the interaction directly and the driver doesn't at least have the ability to override the decision.

As already mentioned in chapter two, it is still questionable whether an extensive scaling of level 3 and 4 vehicles is even possible under the current legal and technical circumstances (Ebers, 2021; Greger, 2018; Othman, 2021). Complication matter even further, there is only little reciprocal understanding between the human and machines (Koo et al., 2016; Krueger et al., 2016; Waschl et al., 2019). This in turn makes it difficult to develop trust and acceptance, since it is not apparent to the passenger on what basis the vehicle makes its decisions (Howard & Dai, 2014; Rudin, 2019). Thus, it is only possible for the passenger to develop trust in the function of the vehicle through experience rather than knowledge (Hoff & Bashir, 2015; J.-H. Lee & Song, 2013). Likewise, the vehicle does not perceive the passenger as a cooperation partner. Possible risks arise from different intentions of the human and machine, which complicate the automation function considerably (M. König & Neumayr, 2017). Therefore, our interest is to connect humans with the machine via communication of the machine to build trust though knowledge. Thus, we adapted the experimental design to investigate whether it is possible to generate trust in a talking self-driving vehicle, even if it is not possible for the potential customers to intervene in the driving action themselves. Therefore, we used the toolkit described here to tests the possible communication strategies of the car in three conditions to examine modulating factors of acceptance.

The experiment described in this section was part of a floating science center of the German Federal Ministry of Education and Research in 2019. Likewise, at the invitation of the Minister of Education, we were able to become part of the Ministry's foyer exhibition, where we installed a copy of the MS Wissenschaft simulator over the summer of 2019. As a result, we could test more than 26,000 test subjects in the virtual world at a low cost and with a mobile data collection setup. We used tracking of head movements and a post-experiment questionnaire to conclude trust and acceptance across the full age range of society. The following chapter will focus on the results of these studies. With these results, we hope to be able to make statements about how a self-driving vehicle should communicate to increase trust and acceptance through specific verbal communication. The results of the study also allow us to get closer to the goal of the thesis, a proposal for a new kind of HMI. The reason for this is that we are able to validate data from the questionnaire with the head movement data of the test subjects. Additionally, we show that it is also possible to draw conclusions about the cognitive state of the passengers. This holds a crucial step for the development of human-machine cooperation.

— 4 —

Talking cars, doubtful users: a population study in virtual reality

Derakhshan S.†, Nezami, F.N.†', Wächter, M.A.†, Keshava, A.,Vidal De Palol, M., Lukanov, H., Czeszumski, A., Pipa, G., König, P. *Talking cars, doubtful users - a population study in virtual reality* IEEE Transactions on Human-Machine Systems., vol. 52, no. 4, pp. 602-612, Aug. 2022, doi: 10.1109/THMS.2022.3168437.

† shared first authorship
' corresponding author

Abstract

ADVs are a significant development in our society, and their acceptance will largely depend on trust. This study investigates strategies to increase trust and acceptance by making the cars' decisions transparent. We created a virtual reality experiment with a self-explaining autonomous car, providing participants with verbal cues about crucial traffic decisions. First, we investigated attitudes toward self-driving cars in 7850 participants by a simplified version of the Technology Acceptance Model questionnaire. Results revealed that female participants show less acceptance than male participants, and there is a general decrease in acceptance with increasing age. A self-explaining car impacts trust and perceived usefulness positively. Surprisingly, it negatively influences the intention to use and perceived ease of use. This implies that trust is dissociated from the other items of the questionnaire. Secondly, we analyzed behavioral data of 26750 participants to investigate the effect of self-explaining systems on head movements during the virtual reality drive. We observed significant differences in head movements during the critical events and the baseline periods of the drive between the three driving conditions. Further, we demonstrated positive correlations between head movement parameters and the TAM scores, where trust showed lowest correlation. This is further evidence for the dissociation of trust from the other TAM factors. These results demonstrate the benefits of combining subjective data obtained by questionnaires with objective behavioral data. Overall, the outcome indicated a partial dissociation of self-stated trust from the intention to use and objective behavioral data.

4.1 | Acceptance model of autonomous vehicles

Introduction

Autonomous driving vehicles are the primary goal of most car manufacturers (Hars, 2016). The development seems to be cumulative since more and more functionalities are automated in new cars (Dajsuren & van den Brand, 2019). One primary reason why ADVs are of value is the possibility of eliminating human driving error, which is responsible for 93% of road accidents (Allahyari et al., 2008; Johnson, 2013). Further, ADVs are safer as they are faster and more precise in the dynamic driving task as well as in object and event detection (Carranza-García et al., 2021; Papadoulis et al., 2019; SAE Internation, 2014; Schoettle, 2017). As technical developments in the field are fast and continuously improving, there is little doubt that self-driving cars will have a significant impacts on our society (Chehri & Mouftah, 2019). These can ranges from drastically decreasing greenhouse gas emissions to reducing the number of traffic-related injuries. Which consequently might lead to possible reshaping the infrastructure of our current cities (Benleulmi & Blecker, 2017; Chehri & Mouftah, 2019; Othman, 2021; Ryan, 2020). Thus, introducing ADVs into our daily lives appears as a highly desirable goal.

Trust and acceptance of potential customers define the extent to which ADVs are used for individual transportation (Howard & Dai, 2014; Krueger et al., 2016). According to current research, there is a limited willingness among potential customers to use ADVs (Kyriakidis et al., 2015; C. Lee et al., 2019; C. Lee et al., 2017; Ryan, 2020; Ward et al., 2017). Various surveys have shown that most potential buyers are unwilling to use an ADV at all or to use it to its full extent (Othman, 2021; Rezaei & Caulfield, 2020). Primary reasons for the lack of trust and acceptance are the fear of system malfunctions and the hesitation of giving complete control to the car (C. Lee et al., 2019; Szikora & Madarász, 2017). The further reluctance of potential customers may stem from low technology self-efficacy (Czaja et al., 2001), meaning that people do not feel confident enough to operate an ADV proficiently (M. König & Neumayr, 2017). Since trust and acceptance are shaped by knowledge and experience, the cause of this reluctance might be rooted in the lack of transparency in and knowledge of the ADVs' decision-making. It might not be clear on what basis an ADV is deciding on. Not knowing what the artificial agent perceives or reasons, directly influences the concerns of safety (Forster et al., 2017; Koo et al., 2016). Therefore it is crucial to find measures that are able to increase the trust and acceptance of ADVs.

According to the technology acceptance model (TAM), perceived usefulness and ease of use are cognitive responses to new technology and predict the intention of using it (Davis & Venkatesh, 1996). Consequently, a low intention to use makes the future application of ADVs questionable (Bergmann et al., 2018; Howard & Dai, 2014; Rezaei & Caulfield, 2020). Belanche and colleagues developed a research model (Belanche et al., 2012) expanding the TAM by adding trust as a component. They found a causal relationship between trust and all three elements of the original TAM (Belanche et al., 2012). Therefore, trust can be seen as a critical factor for the acceptance of a new technology (J.-G. Lee et al., 2015; Lüders & Brandtzæg, 2017; Wintersberger & Riener, 2016). Earlier studies investigated trust modulating factors, such as the human-machine communication style, feedback, and anthropomorphic features in automation (Hoff & Bashir, 2015; Seppelt & Lee, 2019; Wintersberger et al., 2019; Wintersberger et al., 2020). Hoff and Bashir suggested that trust in an ADV is an accumulation of personal tendencies, environment, and user's perception of the autonomous system (Hoff & Bashir, 2015). Lee and See (C. Lee et al., 2019) argue that the perceived homogeneity of communication style, feedback, and anthropomorphic features shape trust levels. The shared statement in all these findings is that the user should perceive the system as reliable and trustworthy. Moreover, previous research showed that excess information about the driving state of a car is perceived as distracting or unpleasant (Howard & Dai, 2014; Koo et al., 2015; Krueger et al., 2016; Othman, 2021; Rezaei & Caulfield, 2020; Ryan, 2020). The desired amount of information by the ADV may be the key to understanding trust and, consequently, acceptance of self-driving cars (Du et al., 2019).

Research on trust as a high-level cognitive phenomenon relies heavily on self-reported data. A review of Raats and colleagues of 258 experiments on trust in ADVs revealed that 84% used questionnaires as the assessment method. Only 4.7% of the studies used observations as a data-gathering tool (Raats et al., 2020). An objective form of data is needed since the participant's self-assessment is often biased by self-perception or socially desired behavior (B. C. Choi & Pak, 2005). For this, we propose head movement data, since humans represent their cognitive states implicitly based on body language, facial expression, gaze direction, and movement of the head (Newen et al., 2018; Zhao et al., 2013). Even though previous research has established the link between gaze shift and cognition (Yarbus, 1967), several studies showed that head rotation corresponds to the visual gaze (Fang et al., 2015; Yarbus, 1967) and both coordinate cognitive processes (Land, 2004; Proudlock et al., 2003). This coordination exists in orientation, which means that the head and eyes move in the same direction. Thereupon, the head orientation provides information about the center of attention (Fang et al., 2015). Behavioral data such as head movements are unconscious, fine-grained, and provide continuous information that can be used to access cognitive processes like trust (Grafsgaard et al., 2012; Lu & Sarter, 2019).

To gain insight into modulating factors of acceptance towards ADVs and their representation in users' head movements, we used a previously developed large-scale virtual reality experiment called Westdrive (Nezami et al., 2020). We expected to find significant differences in the participants' attitudes and significant differences between different age groups and genders. Additionally, we predicted differences in head movements between different conditions. We expected to find significant variance in head movement patterns and head angular velocity as an effect of transparent communication of an ADV. We assumed that the self-reported acceptance in conjunction with head movements enables more objective insights into modulating factor of acceptance.

Materials and Methods

We gathered data from visitors in the German Ministry of Education over six months, and in a traveling exhibition (MS-Wissenschaft) over the course of a full summer. Participants experienced a 90-second drive in a virtual environment called Westdrive, covering roughly 2,5 km^2 with more than 100 cars and 150 pedestrians. Participants experienced a single trial in one of three driving conditions. The first condition was a fully autonomous car with an anthropomorphic voice assistant system (AVAS) giving information about critical traffic events and the corresponding decisions of the car. The second condition was an ADV with a radio broadcast playing through the whole trial. In the third condition, a female Taxi-Driver drove the participant through the city. Here, the TaxiDriver responded verbally to the surrounding traffic. We gathered objective and subjective data in the form of head orientation and head angular velocity, as well as by an adaptation of the TAM questionnaire (Davis & Venkatesh, 1996).

During each trial, participants were confronted with three critical traffic events without the possibility to intervene (Figure 4.1). The duration of the events was the time between entering and exiting the event objects to the participants' view. In the first event, a jogger crossed the road directly in front of the car. In the second event, a high-speed car took the right of way at an intersection and the third event included a slow pedestrian crossing the street. The onset and end of the events were the same for all the participants in all conditions. At none of the events, the participants' cars hit any event objects. In the AVAS condition, the ADV gave short information about the critical event situation. This happened at the spawning of the critical traffic objects to warn at the earliest possible point. The design of these events was based on previous research that showed feedback should include the reason why an ADV decides in a specific way (Koo et al., 2015). Also, additional information should be provided while interacting with vulnerable road users when their intentions are not clear but can influence the car's behavior (Wintersberger et al., 2020). These events were implemented to test the participants reactions as a passive passenger in critical situations. They were designed to test whether the in-vehicle communication can alter behavioral reactions and acceptance.

Figure 4.1: Three scripted critical events occurred during the ride from top to bottom: Pedestrians running on the street from left to right, fast cars cutting in the self-driving car path, and pedestrians walking in the middle of the road.

The simplified questionnaire consists of three questions from the original TAM in perceived usefulness, ease of use, the intention of use, and one additional question on perceived trust. It also included questions on age, gender, aviophobia, driving experience, amount of gaming hours per week and the number of exposures to virtual reality before the experiment. The questionnaire has been answered in the Likert scale, with numbers from 0 (strongly disagree / dislike) to 100 (strongly agree / like) indicated by thumb icons of like and dislike.

The used experimental setup consists of two HTC Vive pro HMDs and lighthouses version 1 for tracking head position and rotation while seated in the car. The VR computers were equipped with Nvidia Geforce RTX 2080Ti GPUs, 16Gb of RAM, and

Intel Xeon W-2133 CPU @ 3.60Ghz core, resulting in an average frame rate of 25,2 fps. Additionally, the setup used two raspberry pies and touch monitors for web-based questionnaires. For analyses, Python 3.6, pandas 0.24.2, NumPy 1.16.4, Scipy 1.7.2, statsmodels 0.10.0, as well as SPSS 29 were used. All plots were created using Matplotlib 3.1.0 combined with seaborn 0.9.0. Data-driven prepossessing on questionnaire data was performed with the OPTBIN algorithm(Knuth, 2013) using histogram-based age binning.

Analysis of the data

Head movement data were obtained from 26750 participants and the questionnaire was answered by a fraction of them. Elimination of incomplete answers resulted in 7850 data sets.

First, we focused on the analysis of the questionnaire data. Of the complete data set, 4464 participants identified as male, 3386 as female. By using optimal binning (Knuth, 2013), we divided participants into five age groups. The cleaned data set consisted of 2812, 1513, 1883, 582, and 86 in the age groups <20 years, 21-40 years, 41-60 years, 61y-80 years, and more than 81 years, respectively. In the AVAS, TaxiDriver and RadioTalk condition we recorded 2691, 2636, and 2509 data sets, respectively. The large number of participants in each bin allowed the use of regression-like inferential tests (i. e. MANOVA) due to their robustness against non-normalities in large data-sets (Pek et al., 2018).

To investigate the effect of gender, age, and driving condition on the four aspects of the questionnaire, a one-way multivariate analysis of variance (MANOVA) has been performed. MANOVA tests the optimal linear combination of dependent variables to find significant effects. We performed a one-way MANOVA for the four TAM aspects modeled with respect to gender, age, and driving condition. Pillai's Trace test statistic uses estimated F-Values to test significance, which is robust against non-normalities. Therefore, Pillai's Trace adds an extra layer of protection against false positives (Finch & French, 2013) and is a good choice for interpreting the results. To understand how different categories within each factor, e.g., male or female in gender, affect the four TAM aspects, we calculated a separate one-way analysis of variance (ANOVA). Following, we calculated the different effect sizes (Cohen's D and Hedge's G) for each of the factors using estimated means and standard deviations reported for the category within that factor. Although both of these effect sizes are based on Cohen's suggestions, Hedge's G considers the sample sizes of the compared groups. Therefore, both effect sizes have been used to interpret the results. Further, each participant's four TAM aspects were combined into one single value. Together with the MANOVA, we were able to make statements about how gender, age, and the condition affect the questionnaire scores.

However, ANOVA can only be calculated on a single independent variable. The best way to combine the four TAM aspects into one value is by multiplying each aspect's score for a given participant by a corresponding weight and adding them all together to get a single value. This acceptance score was calculated by performing a linear discriminant function for each factor that will yielded in a different raw coefficient for each TAM aspect concerning the given factor. The linear discriminant analysis (LDA) intends to find a linear combination of features that characterizes or separates two or more classes. It expresses the dependent variable as a linear combination of the independent variables that maximizes the group differences within the dependent variable (McLachlan, 2005). The raw discriminant function coefficients can be used as weights to calculate the four TAM aspects into one independent number, which we can call acceptance score.

Next, we turn to the analysis of the objective behavior. A head-mounted HMD measured the orientation and position of the participant's head in the virtual environment. We determine the head orientation in a reference frame fixed to the car. Since most interesting visual detail was placed near the ground level and all the dynamic objects of the virtual city moved along the horizontal axis, we focused on the orientation along the horizontal plane. Further, we compared the head angular velocity, meaning the change in head orientation degree over time. To examine the differences between conditions, we used one-way ANOVA followed by the Tukey honest significant difference (HSD) post-hoc test. The Tukey HSD compares pairs of means to detect which of the group means are different from the others (Mean-diff). With this test, we could define the separate condition that causes differences in orientation and angular velocity in specific point of time (Abdi & Williams, 2010). Additionally, we calculated the Pearson Correlation between the head angular velocity and the TAM scores for each questionnaire item to check for consistencies in both subjective and objective measures.

4.2 | Acceptance modulating factors

4.2.1 | Questionnaire Results

The questionnaire data of the simplified TAM from 7850 participants showed a positive correlation of $r>0.4$ between the questionnaire items. Therefore, these items have to be analyzed together as multivariate dependent variables. To check validity of the assumptions, a Levene's test was performed. If the test was significant we would assume a violation of variance homogeneity in the groups. Levenes's test resulted in F-values of 1.369 for perceived Usefulness ($p = 0.089$), 2.333 for Ease of use ($p < 0.001$), 1.459 for Intention of use ($p = 0.053$) and 1.443 for Trust ($p = 0.058$). Considering the large sample size, known to reduce p-values in Levene's test, a further check of the covariance matrices for the dependent variables of the TAM concerning the main factors of gender, age group, and condition has been done. We found homogeneity of covariances, as assessed by Box's test ($p > .001$). Together, Levene's test and the covariance matrices provide essential evidence for the validity of the assumptions of MANOVA. Out of the four different null hypothesis tests of the multivariate analysis, Pillai's Trace was chosen due to its known robustness toward nonnormalities in the data (Ateş et al., 2019). Therefore, the multivariate analysis of variance is the prime analysis method (Warne, 2014).

To gain deeper insights into how gender, age, and condition affect the TAM factors, LDA was used to extract each independent variable's weighted influence. Linear discriminant analysis tries to find a set of coefficients that will maximize the separability within the given independent variable. These coefficients were used to interpret the influence of each independent variable on each of the modulator factors of the TAM.

The effect of gender

First analysis checked for differences between male and female participants regarding the acceptance scores. In order to find out the influence of gender on acceptance, we performed an MANOVA with a follow-up LDA for gender. The Pillai's Trace resulted in 0.00293 ($F(4,7835) = 4.761$, $p < 0.001$) showing that there is a significant effect of gender on overall acceptance (Figure 4.2).

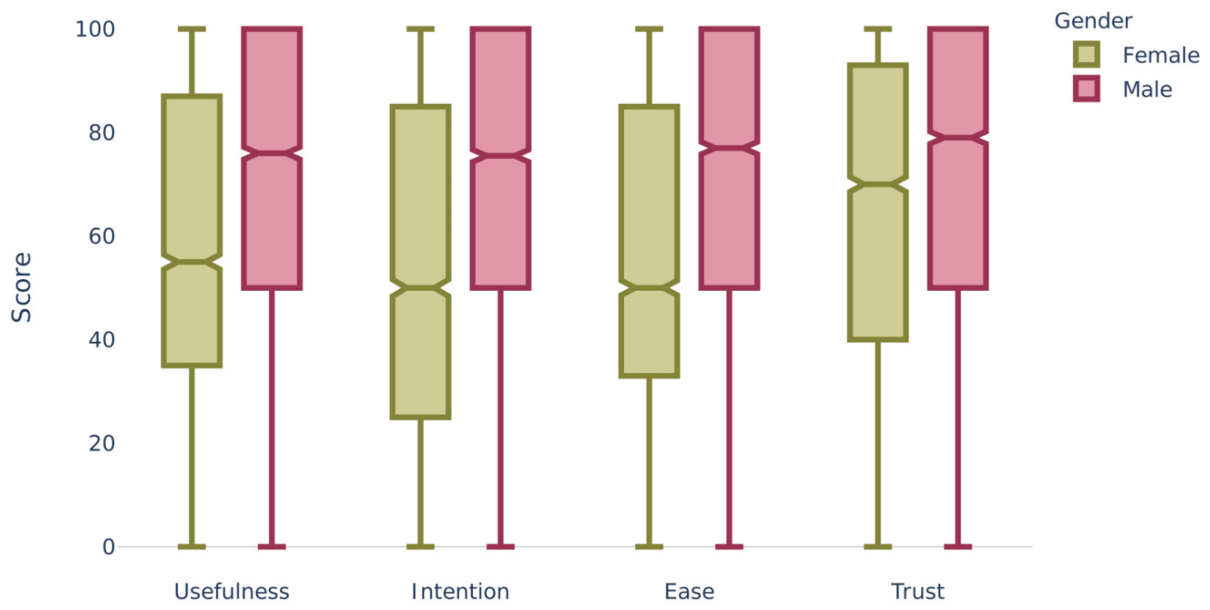


Figure 4.2: The descriptive categorical plot of the mean questionnaire answers for gender.

The follow-up LDA showed that females have a lower score based on the observed discriminant coefficients. The resulting coefficients were -0.33 for the intention of use, -0.06 for perceived usefulness, -0.60 for perceived ease of use, and -0.18 for trust (Figure 4.4 a), all with a medium effect size (Cohen's D = 0.45). Additionally, the LDA showed that perceived usefulness and trust were less affected by gender than the intention of use and the perceived ease of use (Figure 4.3 a). These findings indicate that females and males have an almost equivalent attitude towards the perceived usefulness but differ in the perception of ease of use and, consequently, the intention of using self-driving cars. Thus, we interpreted that females anticipate difficulties in handling and therefore score lower in the intention to use.

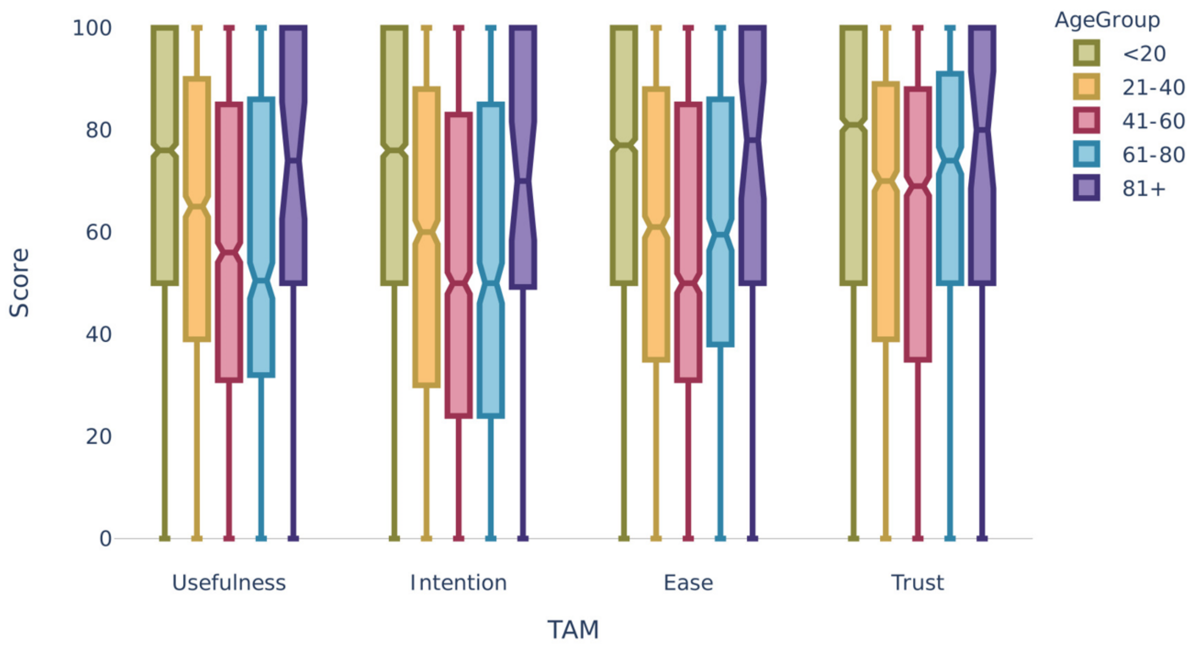


Figure 4.3: The descriptive categorical plot of the mean questionnaire answers for age group.

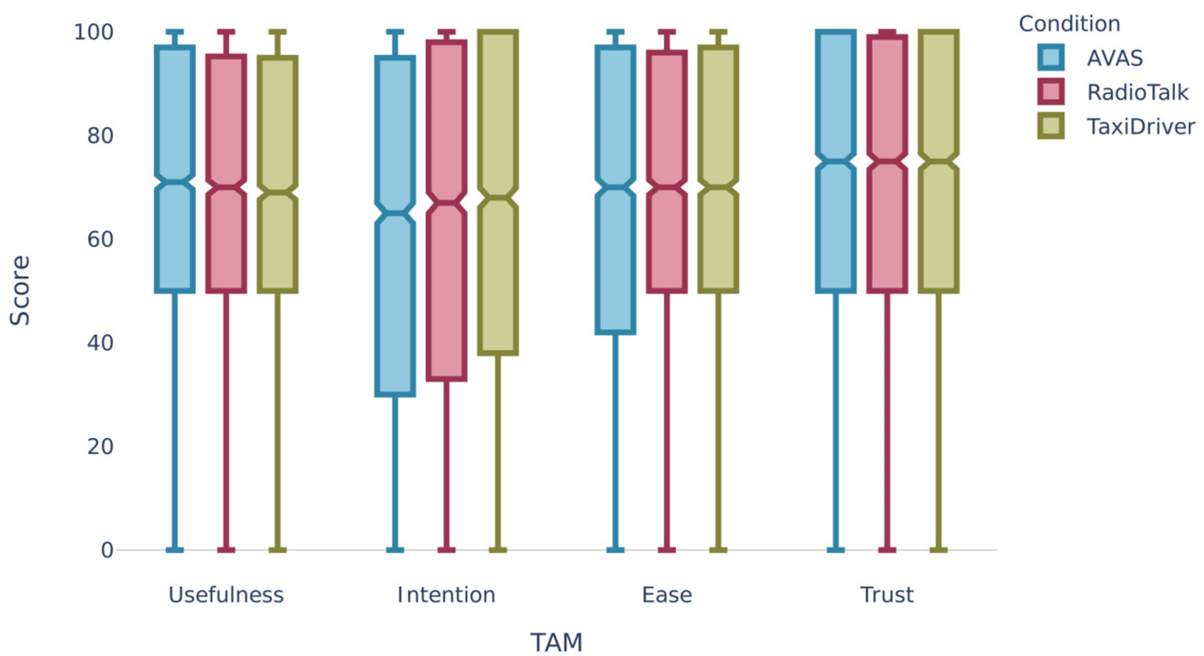


Figure 4.4: The descriptive categorical plot of the mean questionnaire answers for condition.

The Interaction effect of gender and age group

While investigating the effects of gender, age and condition, it became clear that these factors separately did not explain all variance observed in the data. There was a significant interaction effect of gender and age group with Pillai's Trace of 0.00498 ($F(16,31352) = 2.441$, $p = 0.001$). According to the follow-up LDA, there was a negative effect for the intention of use and perceived ease of use (both -0.73) and a positive effect on the perceived usefulness (0.22) and trust (0.55) in the questionnaire items. Here, the effect sizes were most notably between the age groups 21-60 years compared to under 20 years and above 60 for each gender. These results support findings of the previous analyses on gender and age. In addition, it could be shown that the interaction of gender and age had a significant influence on the acceptance of ADVs. In addition, the largest effect sizes (0.5 < Cohens' D <= 0.9) resulted by comparing female participants in the age between 21-60 years against the male participants in the same range. Participants below 20 years had the highest TAM scores, and females between the ages of 21-60 years showed the lowest TAM scores. Although there is was a decrease in all TAM factors in both genders for ages between 21-60 years old compared to the below 20 years, female participants showed stronger decreases in TAM scores (Figure 4.5). This accounts especially for the intention of use and perceived ease of use. Once again, as age increases for people between ages 21-80 years, we can also observe TAM scores. In conclusion, gender and age group interaction significantly affect all TAM factors, specifically negative influences on the intention of use and perceived ease of use for ADVs, but a positive effect on perceived usefulness and trust. This means that although the ADV was seen as useful and trustworthy, there were still other hidden factors that decreased the ease of use and the intention to use it. Consequently, the demographic factors of age, gender and the interaction of these two, have much more impact on the items of the TAM questionnaire. The positive effects of a self-explanatory ADV were not sufficient to compensate for the negative influence of demographics on ease of use and, accordingly, intention to use.

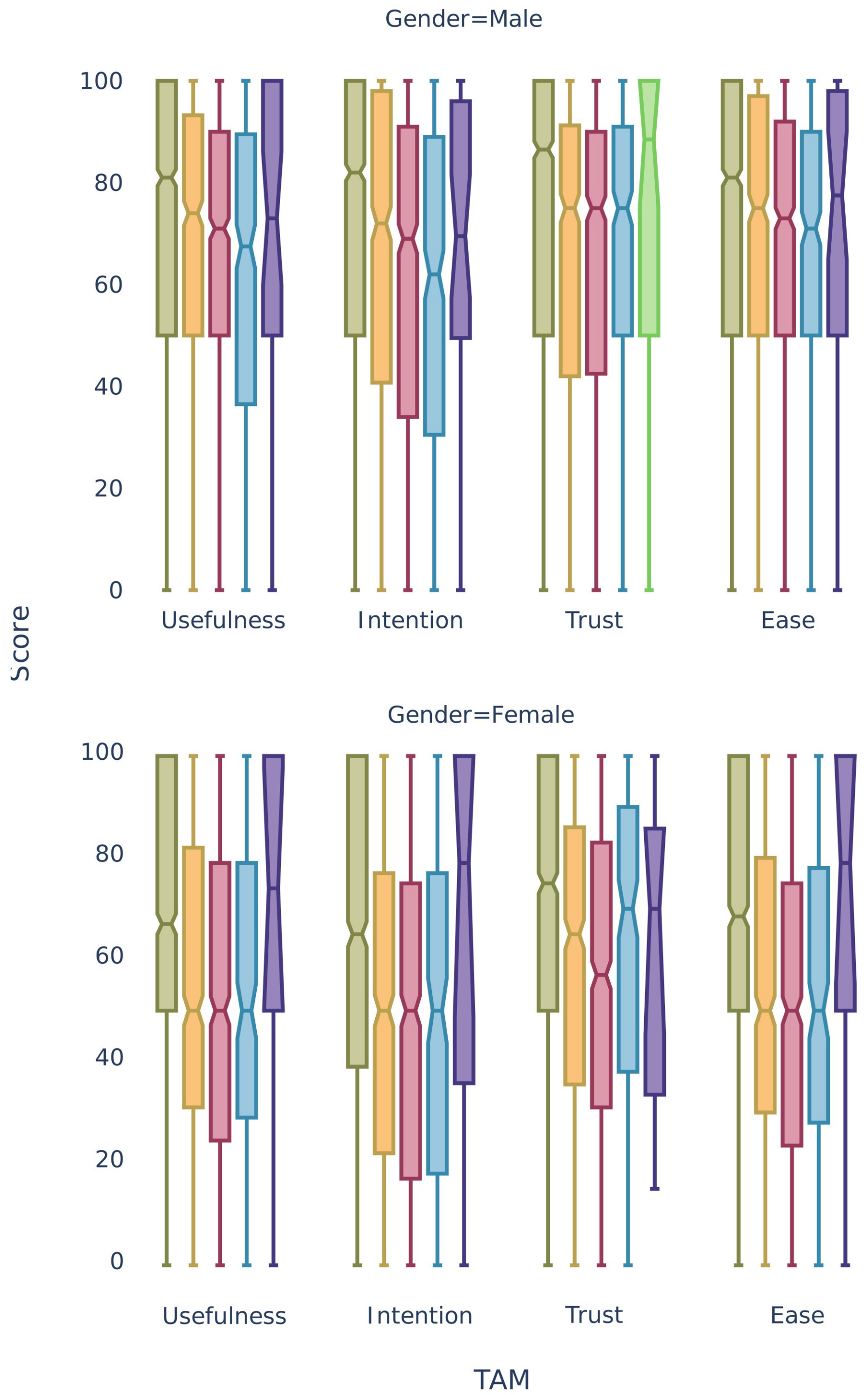


Figure 4.5: Mean of answers for questions regarding the usefulness, intention, trust, and ease for age group and gender combined

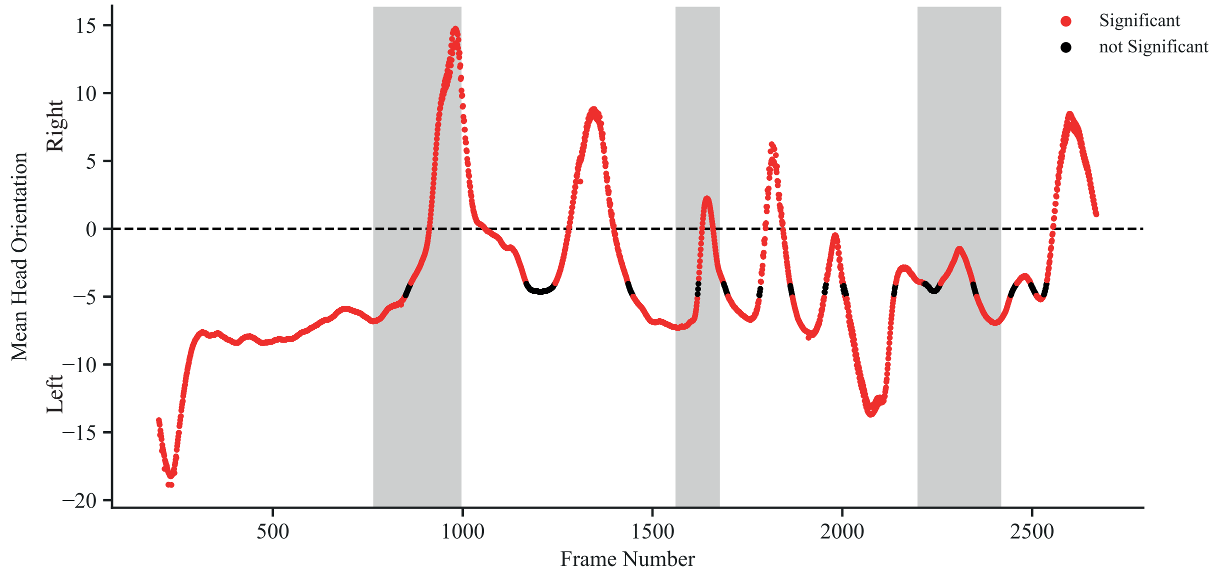
4.2.2 | Behavioral Results

Identification of critical events

As a first step of the head movement analyses, we investigated whether participants' behavior differs during the critical events from the baseline parts of the drive. The initial analysis was performed regardless of the driving condition. We considered the mean orientation and variance of head orientation over all participants as the relevant dependent variables. Collapsing the data over conditions, we tested whether the mean of head orientation in each frame was significantly different from the distribution resulting from a permutation over time (permutation test). Head orientation differed significantly from baseline at the end on the first and second and at the very end of the third event. (Figure 4.6). Further, we observed differences early in the trial, when participants where intensively looking around inside the car. Additionally, three other significant intervals were observed. During these times pedestrians were visible on the sidewalk in crowded places of the city. We assume that this is related to a need of information to assess the situation. A final period of deviant head orientation is observed at the very end of the drive, when participants prepared to exit the car in a crowded area. By applying this method we are confident that an additional measurement of head movements is a valid approach to enhance subjective data. Overall, compared to the baseline head orientation, the three critical events showed significant differences in participant behavior regardless of the effect of conditions. These differences were not limited to the critical event intervals but identified in additional areas of the trial.

The effect of condition

As a next step, we consider how much of the observed variance was related to the effect of condition. We investigated whether participants' behavior objectively differs between conditions. Differences in head orientation were seen as indicators of participants' reaction to the environment in different conditions. By visualizing the head movement data, we observed differences in the mean head orientation, over large parts of the drive and during the critical traffic events. (Figure 4.7). To see, whether these differences were significant, we calculated a one-way ANOVA based on head orientation for each frame as the dependent variable and applied a post-hoc comparison of Tukey HSD in the significant Intervals. The result of the ANOVA showed significant differences in head orientation between the three conditions during most of the drive time ($F>10$, $p < 0.05$) (Figure 4.8). Specifically, the TaxiDriver condition was significantly different from the two others over a large part of the drive. The post-hoc comparison revealed larger mean difference (Meandiff) a)



b)

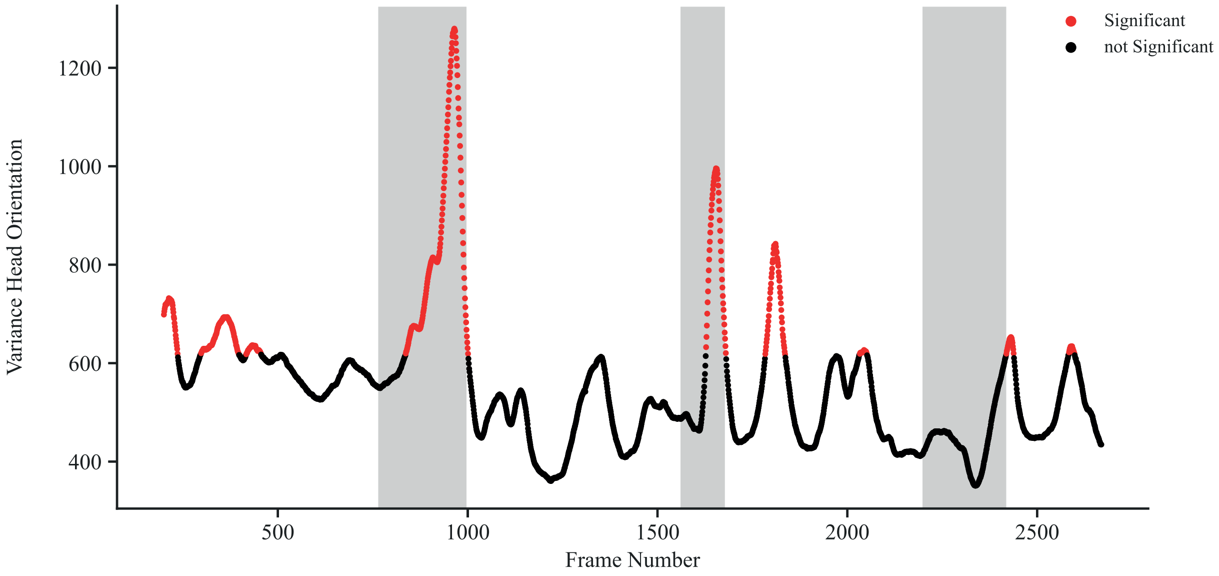


Figure 4.6: Time intervals of significant differences in head orientation after the permutation test (n=1000). Shaded areas represent the critical traffic event intervals. a) Mean of head orientation over all subjects. Each point indicates an average of the head orientation across the participant within each frame. b)Variance of head orientation over all subject. The red areas indicate the intervals where there was a difference in head orientation between all participants. Please note that the data is collapsed over condition.

for TaxiDriver compared to AVAS (i.e. Meandiff = 2.07, p = 0.001 for frame = 1300) and RadioTalk (i.e. Meandiff = 1.77, p = 0.001 for frame = 1300) condition. This result was mostly constant during the experimental trial, including the three critical events. At the start of the first and the second event, no significant differences between the RadioTalk and AVAS condition were found. In the third event we found differences between all three conditions. Here, participants elicited the smallest degrees of the head orientation in the AVAS condition, and the highest degree in the TaxiDriver condition. We mainly observed a higher mean head orientation in the TaxiDriver condition in the beginning of the critical traffic events. This means, the distribution of the head orientations in angular space were wider in the TaxiDriver condition compared to the two autonomous conditions in most of the trial times (Figure 4.7)

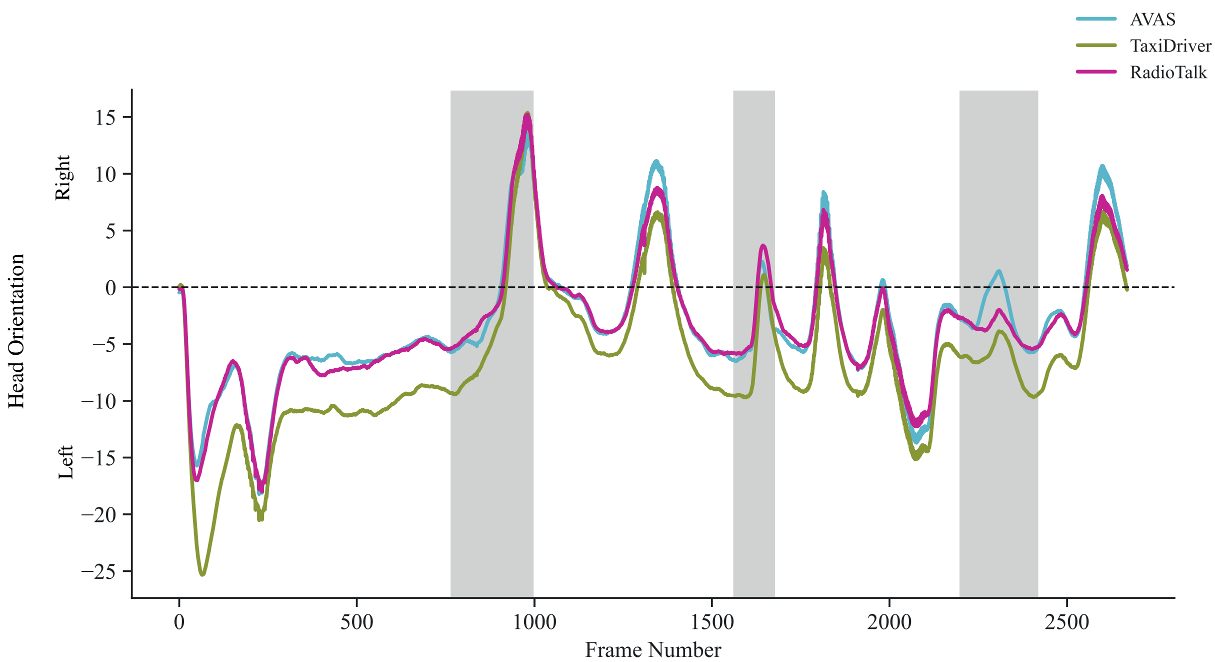


Figure 4.7: Mean head orientation in each frame divided in three conditions. The positive and negative values of the mean orientation relate respectively to the right and left directions. Shaded areas represent the critical traffic event intervals.

Head angular velocity

To gain deeper insights into the participants' head movement behavior, we calculated the magnitude of change in head orientation throughout time as the angular velocity. We quantify the absolute value of the angular velocity for each critical traffic event separately, based on the experiment's overall average frame rate. The analysis of the angular velocity showed that in the AVAS condition, participants rotated their heads significantly faster only in the first critical traffic event ($F(2,24447) = 71.35$, $p < 0.01$). In the second critical traffic event, no significant differences between conditions were found ($F(2,24447) = 2.8$, $p= 0.06$). In the third critical traffic event, the angular velocity in AVAS was significantly lower than the two other con-

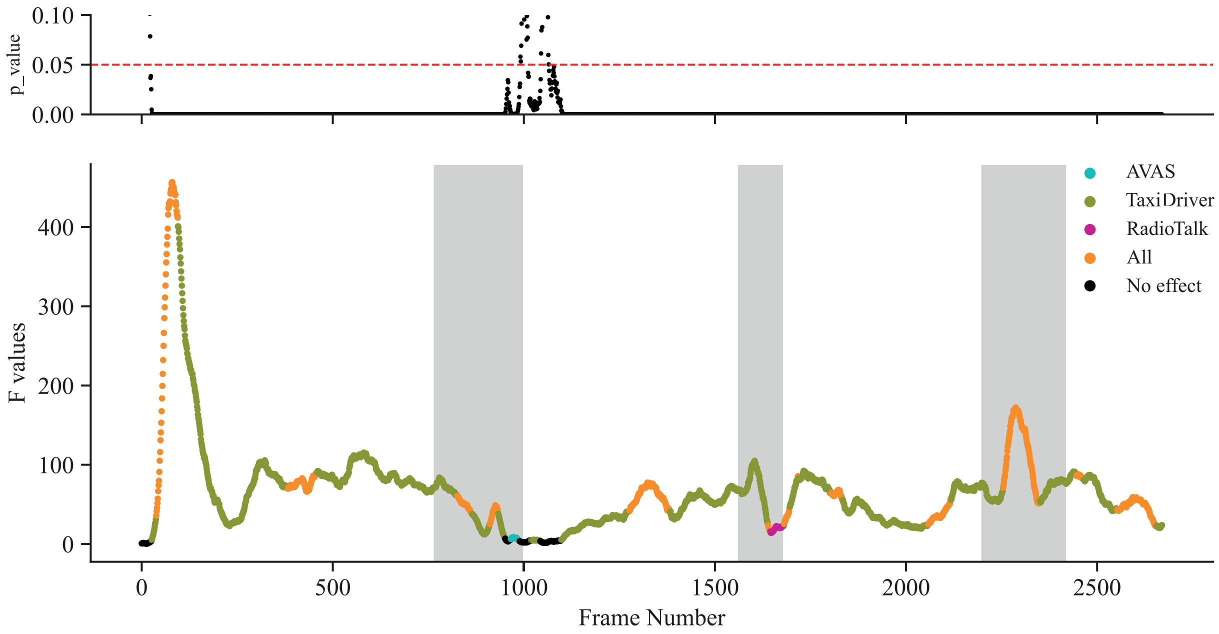


Figure 4.8: Time intervals of significantly different behavior between the three conditions. The graph depicts the P and F values of one-way ANOVA overall the experimental trial. Each dot shows the original F value of each frame. The red dash indicates the significance threshold ($p < 0.05$). Shaded areas represent the critical traffic event intervals. The result of Tukey's post hoc comparison is represented by different colors. Each color shows the significant variable mean(s) in cross-check.

ditions ($F(2,24447) = 29.06$, $p < 0.01$) (Figure 4.9). Overall, the data revealed that the angular velocity of head movements decreased during the experimental trial in all three conditions. However, the AVAS condition reduced the head's angular velocity to a larger degree than the other autonomous condition. With the analysis of the angular velocity, we were able to show that participants' behavior changed as an effect of self-explaining ADV over time.

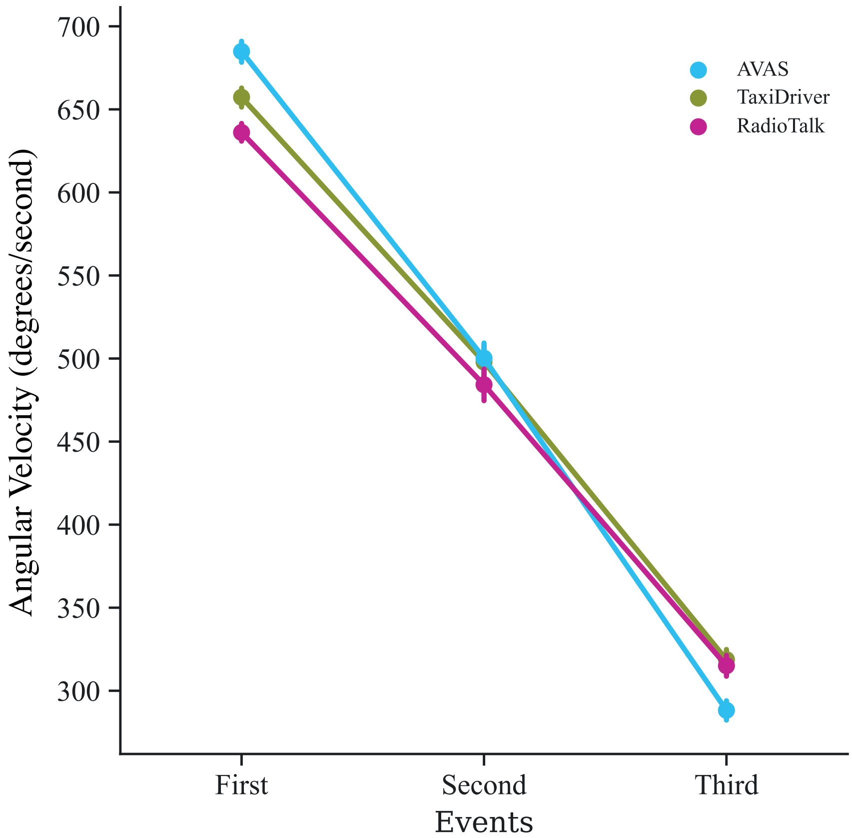


Figure 4.9: Head angular velocity for the critical event intervals. The y axis refers the rotation change divided by the number of event frames.

Questionnaire comparison

The head angular velocity was an illustration of the participant's head movements behavior during the trial. Calculating the relationship between the angular velocity and the TAM items allowed us to determine if the participants self-assessed acceptance had been expressed in their previous behavior during the experimental trial. We used Pearsons' correlation for the participant's absolute head angular velocity over the entire trial and participants' respective TAM item scores. The analysis showed positive correlation between the head angular velocity and all TAM item scores in all three conditions (Figure 4.10). Comparing the TAM items, there was a lower correlation between the angular velocity and trust compared to its correlation with other TAM items. Along with the previous finding in the analysis of the questionnaire, the mismatch between trust and the other questionnaire items has been demonstrated in the correlation of the items and the angular velocity. The dissociation between trust and the other questionnaire items lead us to the assumption that trust is not an ideal item in self assessments via a questionnaire. This claim is supported by the mismatch in the self-assessment, as well object behavioral data. Therefore, we argue that the objective behavioral data was able to reflect the findings in the TAM questionnaire.

Figure 4.10: Summary of the Pearson correlation between the head angular velocity and the TAM questionnaire items. The correlation p value for Intention, usefulness, ease of use and trust are as follows for each condition: a) AVAS (Intention: p <0.001, Usefulness: p <0.0010, Ease of use: p <0.001, Trust: p <0.01) b) RadioTalk (Intention: p <0.001, Usefulness: p <0.001, Ease of use: p <0.001, Trust: p <0.001) c) TaxiDriver (Intention: p <0.001, Usefulness: p <0.001, Ease of use: p <0.001, Trust: p <0.001)

4.3 | Implications for future in-vehicle communication

The present study revealed that self-reported acceptance in conjunction with objective observation enhances the understanding of modulating acceptance factors. According to the results, subjective data from a post-experimental questionnaire and objective data from head movements during the experimental trial were largely congruent. Outcomes of investigating gender, age, and condition effect on the overall acceptance showed a lower acceptance of female participants toward ADV than males. However, this effect is even more pronounced in the intention to use ADVs. The results also indicated that people below 20 years of age have the highest acceptance toward ADV, gradually decreasing with age while increasing again above 80 years.Regarding the effect of a self-explaining ADV, we found a small positive effect in the ease of use and a small negative effect regarding the intention of use. However, age, gender, and the interaction of these two have a substantially higher impact on the questionnaire scores.

Therefore, the positive effects of a self-explanatory ADV are not sufficient to compensate for the negative influence of demographics on ease of use and the intention to use. We could show that participants' head orientation differed between the conditions by analyzing the head movement data. Especially in the TaxiDriver condition, we did see significant differences over the whole drive with further differentiation between conditions during the critical events. Further, we observed a decrease in the head's angular velocity over time for all conditions. This effect was most substantial in the AVAS condition. Finally, correlating the magnitude of the participant's head angular velocity with the TAM scores showed a significant relationship between acceptance as a combination of the TAM factors, which was weaker for the factor trust.

Previous studies mainly depended on answers gathered from the potential users (Howard & Dai, 2014; Raats et al., 2020; Wintersberger et al., 2020). However, behavioral data is not as susceptible as questionnaire answers and can be used to validate possible self-assessments (Davis & Venkatesh, 1996). The presented study could show a dissociation of the self-assessed trust and the other TAM items, especially with the intention to use. This observation contrasts previous research such as the work of Belanche (Belanche et al., 2012). In fact in the present study it is shown that self-assessments are heavily modulated by the demographic factors such as age and gender, as well as the interaction of these two factors. Behavioral data confirmed the dissociation of trust and intention, by showing a connection between head movements and scores in intention, ease of use and perceived usefulness.

Therefore we are arguing that including behavioral data is a valid approach to better understand underlying factors of acceptance and correcting potentially faulty subjective data. This is due to the fact that head movements can be considered as part of nonverbal communication in humans (Mehrabian, 2017). It contains information of the participant's emotions and intentions (Gunes & Pantic, 2010). For instance, head angular velocity and acceleration were higher during negative affects (Hammal et al., 2015). Combining sources of subjective and objective data, make it is possible to validate the questionnaire data. In conclusion, it could be stated that the behavioral data is an important resource that can be used to validate investigations of the technology acceptance model and its underlying factors.

Due to the nature of the experiment within a public exhibition and a significant number of visitors, the technology acceptance questionnaire used in this study was a simplified version. Thus it might not grasp the entire aspect and spectrum of factors that modulate acceptance, such as technology self-efficacy, which might play an essential role in perceived ease of use. Furthermore, the questionnaire was also translated into German, and we could not validate the questionnaire before using it in the experiment. Although part of the variance in the data might be due to the translation, such effect is thought to be minuscule and negligible since our main findings align with that of the previous works (K. Chen & Chan, 2011; Koo et al., 2016; Othman, 2021; Venkatesh, 2000). Due to the simplified nature of the study, we can not directly address and analyze the underlying information processing that influences attitude. Still, we are confident to make informed statements due to the large effects in a vast data set. Additionally, there is a possibility that cybersickness influenced the TAM scores and head movement data.

Nevertheless, we tried to control as much as possible for motion or cybersickness in this trial. The first step was to include a bigger static frame for the participants as the car interior, reducing the probability of cybersickness during the trial. Also, we only used a low-speed environment, without sharp turns to reduce cybersickness as much as possible (van Emmerik et al., 2011). Further, We acknowledge that a more precise measurement instrument such as eye trackers would have enhanced the analysis and the findings. However, once more, the nature of the experiment and the absence of experimenters on-site (not counting the numerous visits for maintenance) rendered the use of such methods impossible. Another criticism could be that the experimental time was limited to 90 seconds, and each participant observed only one of the experimental conditions. However, this experience can already investigate participants' acceptance toward various in-car communications in ADVs. Additionally, the vast amount of data gathered by the experiment allowed for entirely data-driven analyses both for questionnaire and behavioral data. Therefore, the results of this study are valuable for understanding the population's acceptance of ADVs and the importance of objective measurements.

Despite these limitations, we are confident to show an effect of a self-explaining ADV based on subjective and objective data. As mentioned earlier, previous research explained trust as a combination of the communication style, feedback, and the anthropomorphic features of the ADV (Belanche et al., 2012; Koo et al., 2015). In contrast, Hoff and Bashir state that trust is largely shaped by the personal traits of the users (Hoff & Bashir, 2015). This is supported by newer findings in real driving scenarios, where personality traits were identified as relevant factors of trust (Stephan, 2019), and were only out weighted by the actual driving performance. These factors are summarized under "dispositional trust," which consists of age and gender, and personality traits. In line with previous research our study could show that the demographic factors have a higher impact on acceptance compared to the ADVs' features.

Nevertheless, our findings are not generalizable over all demographic groups, since their communication needs are different: While we see a positive influence of the talking car in one group, the second group may view the in-vehicle information as excess and distracting. Following, the user-specific communication could increase trust in doubtful users - making them more confident to properly operate such a system since it might be able to increase system knowledge. However, there is a need for a further investigations using more extensive questionnaires to examine further modulators of acceptance, specifically trust in combination with more objective measurement instruments such as eye tracking. In the end, we argue that user specific in-vehicle communication can be useful to create guidelines for the further development of a safer and inclusive future of mobility.

Acknowledgment

The authors thank all involved members of the Westdrive team that helped building software and hardware as well as acquiring data. Namely, we would like to thank: Johannes Max Pingel and Sumin Kim for helping to organize and commenting the experimental code, again Johannes Max Pingel for maintaining hardware during the duration of the exhibition. Deniz Gün and Thomas Klein for implementing the supplementary material into the exhibitions, as well as Philipp Spaniol for designing and building the exhibitions and lastly Marc Vidal de Palol and Hristofor Lukanov for implementing the Questionnaire, supplementary material, and server structure. Without the support of all these wonderful members of the institute, this project would not have been realized.

Author Contributions

SD is the leading author and took care of the data analyses in this manuscript. FNN and MAW were developing and designing and conducting the experiment. MAW assisted in the manuscript writing process. FNN and AC assisted in the data analyses and manuscript review. PK and GP supervised the project.

Datasets

Code of the entire experiment conducted in this article is available in project-westdrive Gitlab repository under creative common license. furthermore all analysis scripts and output resuts, as well as raw data and demo video are available at OSF under creative common license.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Chapter Summary

In this chapter, we have seen that in-vehicle communication influences attitudes towards an automated vehicle. We were able to show that a vehicle of automation level 4 generates more trust if it has anthropomorphic features and informs the occupant about what is currently happening in traffic. Thus, we were able to replicate the results of previously conducted research stating that information given by the car increases trust among the passengers (Koo et al., 2015; Wintersberger et al., 2020).

However, we were also able to show that trust does not seem to be the crucial factor for acceptance. Although the vehicle with a verbal feedforward and feedback positively affected perceived usefulness and trust, we saw a negative effects for the intention to use such a vehicle and the perceived ease of use in the data. Comparing the results of this study with our considerations from the introduction, we see that the basic idea of the experiment may have been too naive. In the beginning, the idea was to increase acceptance by providing more information. In the experiment we see that existing doubts cannot be resolved only by verbal communication provided by the vehicle. Although we were able to show an enormous influence of age and gender on the items of the TAM questionnaire, we currently lack actual insight about the questionnaire items that modulate trust. Being currently bound to the gathered data and taking into account that the difference between subjective trust and intention may be due to the simplified form of the TAM, it is also possible that the self-assessment questionnaire is not the most appropriate way to measure acceptance of self-driving vehicles, as expectations and social norms may distort it.

With respect to the question this thesis aims to answer, i.e., how trust and acceptance in artificial agents can be established, a first hint arises here. Based on the study results, a positive effect of verbal communication in critical traffic events does increase the perceived usefulness and trust. Nevertheless, we also found a mismatch between trust and the intention to use such a vehicle. Looking into the possible user groups, there is no generalizable solution to this problem since demographic factors play a more important role in acceptance than the verbal communication strategies of the vehicle with its occupants. Thus, it is the vehicle that should adapt to the respective passenger and their needs.

As briefly described in the discussion of the current paper, we could define three different groups of potential users. The first group contains the "skeptics", predominantly consisting of women between the age of 40 and 60. This group sees the vehicles as practical but not very usable. The second group is characterized by uncertainty. On the one hand, its members clearly see the benefits of the technology and trust it but on the other hand this group responds negatively to the provided information by the car. This is supported by current literature, according to which negative effects on acceptances increase with additional information (Koo et al., 2015; Othman, 2021). This group also includes the group of the oldest participants. Here, we hypothesize that the group of uncertain users perceives the technology as usefulness, since members of this group have an increased need for mobility and a decreased ability to drive by themselves. Nevertheless, they would prefer a traditional form of mobility. The last group can be described as the overestimators. These are the particularly technology-savvy people as well as very young people under the age of 18. According to supporting literature, this group has a distorted perception of benefits and risks of self-driving vehicles. An overestimation of benefits might lead to a misuse of self-driving cars, the disappointment of initial users, and could have fatal consequences (Gillmore & Tenhundfeld, 2020; M. König & Neumayr, 2017). Since these three subgroups have contradictory requirements concerning the in-car communication of self-driving vehicles, it should be possible to satisfy each user groups' individual needs. How exactly an adaptive HMI could be realized and at which automation level which interaction between man and machine will have the greatest benefit still needs to be evaluated by further studies.

Additionally, we were able to show that the different forms of communication in a vehicle influence the explicitly reflected acceptance as well as the behavior of the test subjects. Correlating both data measures we were able to show that subjects exhibited larger but slower movements when they were more accepting and faster movements when they were less accepting. We concluded that we can distinguish between exploration and orientation behavior. Thus, we see that head movement can be a valid instrument for examining implicit attitudes toward self-driving vehicles. Based on the data of the questionnaire as well as the data from the head movements, it is possible to create a representation of the current driver state that, in combination with other data sources, can lead to a tailored in-vehicle communication in the form of an adaptive HMI. This technique could use date from the current environment as well as data from the passenger to adjust it's behavior on the road as well as the communication toward the passenger to increase trust.

More specific, an adaptive HMI allows for a change in communication strategies based on different user groups as well as specific data sources from the passenger. For example, it would be possible for the vehicle to provide less information to passengers during slow and large head movements since this behavior represents trust and exploration of the environment. Hence the negative effects of obsolete information can be avoided. In contrast, tailored information would be provided during faster movements in the direction of travel, since this is thought to correlate with focus on other road users and the assessment of the current traffic situation and therefore with distrust and stress. Here, more information about the current traffic situation could be a possibility to increase the explainability of decisions made by the vehicle, and accordingly, also the acceptance of the human passenger. In such cases, the car would detect when and about which objects the driver needs additional information in order to understand the vehicles actions and to build trust.

Additionally, such a feature in the car would not only increase trust and acceptance, it could be facilitated to help in the OOTLU in a takeover request. Once the car is requesting for a takeover, the car could detect the state of the driver and provide the human with specific information to foster as fast and precise takeover maneuver. One example on how to pair a driver-state specific request could be a combination of the driver-state detection and a audio-visual HUD that was tested in LoopAR. Here the car could detect the informational needs of the passenger via the head movements and guide the head orientation in a takeover scenario through visual input in the windshield to enable a safe and fast take over.

However, there will a small but undeniable probability of scenarios in which humans will not react in a timely and appropriate manner to a takeover request. In such cases, the vehicle must switch into a minimal risk state, that was described further up in the introduction as a vital part of the level 4 and 5 features (SAE Internation, 2014). Ergo, it has to remain in control and be able to come to a safe position even in a state of high uncertainty. Still the question remains, what would this state look like in borderline cases e.g. a total sensor failure, where the car turns blind, or in an unforeseen emergency situation with short braking distances, e.g. an oncoming vehicle on collision course, where a collision will definitively occur. Should we avoid a possibly deadly collision under all circumstances, even if this poses a threat to the vehicle's passengers? Or should the vehicle protect the occupants at all costs? Is a vehicle allowed to swerve and risks injuries to vulnerable road users such as pedestrians or people on bicycles? Should the car stay on the road or is a turn onto pavements a legitimate action in these cases? This raises the question of how people judge such situations as potential consumers and thus enablers of this technology. Is it possible to collect preferences under certain ethic theories or is there no unifying preference? This will now be considered in detail in the next chapter. In this context, test subjects are confronted with a variety of forced-choice dilemma scenarios in which they have to decide which life should be protected and which should be sacrificed. So far, it is clear that there must be support human supervisors in order to automate personal transportation, as they can also perform tasks that are not relevant to driving (Bengler et al., 2020; Endsley et al., 2003; Pohl et al., 2006; Wintersberger et al., 2020). Sooner or later, this will lead to drivers being asked by the vehicle to take over, in which case they will have to react fast and precisely. Since they may be distracted, it can be assumed that these drivers will not react adequately and quickly (Jarosch, Bellem, et al., 2019).

Similarly, we have seen that when supporting humans in highly automated driving scenarios, there is no single communication strategy that generates trust among all possible passengers (Koo et al., 2016; Othman, 2021; Wintersberger et al., 2020). Rather, it is possible to classify users into groups according to their different needs (Belanche et al., 2012; Hoff & Bashir, 2015). In addition, we were able to show that the behavior of the test subjects allows for conclusions about their acceptance in the self-assessment. This knowledge will now be transferred to the next chapter. The goal is to link the previous knowledge with the decisions of the subjects of the moral dilemma scenarios. This would enable us to create a design for a human-centered HMI that can solve design related issues of how a car should interact with its passengers while also solving ethical and legal concerns of automation.

5

Moral Dilemmas in fully Autonomous Vehicles

Faulhaber, A. K.†, Dittmer, A.†, Blind, F.†, Wächter, M. A.†’, Timm, S.†, Sütfeld, L. R., Pipa, G. & König, P. (2019). ***Human decisions in moral dilemmas are largely described by utilitarianism: Virtual car driving study provides guidelines for autonomous driving vehicles.*** Science and engineering ethics, 25(2), 399-418. doi: 10.1007/s11948-018-0020-x

† shared first authorship
’ corresponding author

Abstract

Ethical thought experiments such as the trolley dilemma have been investigated extensively in the past, showing that humans act in utilitarian ways, trying to cause as little overall damage as possible. These trolley dilemmas have gained renewed attention over the past few years, especially due to the necessity of implementing moral decisions in autonomous driving vehicles (ADVs). We conducted a set of experiments in which participants experienced modified trolley dilemmas as drivers in virtual reality environments. Participants had to make decisions between driving in one of two lanes where different obstacles came into view. Eventually, the participants had to decide which of the objects they would crash into. Obstacles included a variety of human-like avatars of different ages and group sizes. Furthermore, the influence of sidewalks as potential safe harbors and a condition implicating self-sacrifice were tested. Results showed that participants, in general, decided in a utilitarian manner, sparing the highest number of avatars possible with a limited influence by the other variables. Derived from these findings, which are in line with the utilitarian approach in moral decision making, it will be argued for an obligatory ethics setting implemented in ADVs.

5.1 | Ethical Implications of Autonomous Vehicles

Introduction

Since their invention in the nineteenth century, cars have considerably influenced townscapes and societies all over the world. Due to the continuous development and increasing sophistication of vehicles, this impact is still ongoing. It even seems that car manufacturers are getting closer to reaching another milestone: a car that is capable of driving without a human driver.

In the last few years, there have been substantial advancements in the development of such Autonomous Driving Vehicles (ADVs). Many features of automation, such as cruise control, camera-based blind spot assistance, and parallel parking have already become standard in modern cars. The majority of car manufacturers and service providers, such as Uber, are currently working on ADVs and planning to commercially market them by 2025 at the latest (Hars, 2016). The fast introduction of ADVs is due to the expected advantages. These might include higher mobility for people unable to drive a car (e.g. elderly, tired, disabled people), better organized traffic and fewer traffic jams due to communication between vehicles. Most importantly, based on improved driving behavior and shorter reaction times the number of traffic accidents and casualties is expected to decrease significantly.

However, with the development of disruptive technologies, new problems arise. Because introducing ADVs might have a large impact on society, critical issues spread over a wide range of areas including psychological, ethical, socioeconomic, and legal aspects.The most pressing issues that need to be addressed include liability in the case of casualties as well as the ADV's behavior in moral dilemma situations. Moral decision-making seems to have little implication for traffic so far because most accidents happen in a split second without the time and the information to think thoroughly about one's reaction. Therefore, humans base their decisions in such situations mostly on reflexes and instincts rather than deep thoughts.

This will change with the introduction of ADVs given that the car's decisions in all kinds of possible traffic scenarios will be programmed beforehand including guidelines for unforeseen events and even highly unlikely scenarios (Lin, 2013, 2015). But there is no consensus yet on who decides what should be programmed. One possibility would be that users may choose an individual ethics setting themselves. However, Gogoll and Müller, 2017 criticized and rejected this option as this would most likely lead to a prisoner's dilemma [1] in traffic (see also "Discussion" section). In their thought experiment, people would choose a sub-optimal and thus a negative outcome for the entire society, just to prevent a possible exploitation by other road users. In consequence, Gogoll and Müller call for an ethics setting that is mandatory for all ADVs. For the implementation of such a setting, an ethical framework is needed which remains widely debated (Hevelke & Nida-Rümelin, 2015a, 2015b, 2015c).

One problem is that people are in favor of ADVs programmed in a utilitarian way but state they would themselves not want to buy such an ADV (Bonnefon et al., 2016). Moral decisions by autonomous systems are often discussed on the basis of trolley dilemmas. The classical trolley dilemma was introduced in 1967 as a philosophical thought experiment (Foot, 1967). The key element is a trolley heading straight toward a group of people, who are on the rails and unable to escape. There is, however, a side track on which a single person stands, unaware of the trolley. Participants in this thought experiment are standing next to a lever that enables the trolley to switch to the side track, resulting in a moral dilemma. Without intervention, the trolley will kill the group of people on the main track. Upon pulling the lever, the trolley will continue on the side track, killing only one person. How do people make decisions in such situations and what moral principles govern their decision process? This question has been investigated and debated extensively (Mikhail, 2007; Thomson, 1976, 1985; Unger, 1996) .So far, research on modified trolley dilemmas in the context of ADVs focuses on whether there is a moral argument for ADVs to act in a deontological or utilitarian way. The distinction between the ethical theory concepts of deontological motivations and utilitarian motivations is hard to draw, especially with a broad notion of deontology. We define utilitarian actions, opposed to random behavior or refusal of behavior, as those maximizing utility by seeking to cause as little overall damage as possible, based on some probabilistic view of the future. This might even include willingness to risk harm for oneself.

Recent studies showed that people in general act in utilitarian ways and are relatively comfortable with utilitarian ADVs, programmed to minimize harm (Bonnefon et al., 2016; Skulmowski et al., 2014). Besides the deontological approach and utili-

[1]The prisoner's dilemma is a mathematical theory based on game theory. Imagine two prisoners accused of committing a crime together. The two prisoners are interrogated and can not communicate with each other. If both deny the crime, both receive a low punishment. If both are confessing both receive a heavy sentence. However, if only one of the two prisoners confesses, he or she leaves the court without a sentence, while the other gets the maximum sentence. The dilemma in this situation is, that every prisoner must choose to either deny or confess without knowing the other prisoner's decision. The sentence depends on how the two prisoners testify together, and thus depends not only on their own decision but also on the decision of the other prisoner.

tarianism, there are many more ethical standpoints concerning how to tackle the problem of decision-making in self-driving cars. These range from virtue ethics, meaning that specialists and ethicists influence the decision-making in machines as a governmental committee, to a machine that mimics the entire spectrum of human behavior (Goodall, 2014). Within these discussions, there is much disagreement in the literature about which ethical setting is the right one to choose and no clear statement could be made yet (Lin, 2015). For the present study, the behavior of the participants served as a starting point.

The intention is to deduce rules from human behavior that would be applicable to all ADVs. This is because people have to agree to an ethical setting that is implemented in their car to actually use it. Moreover, people usually do not judge a case based on deontic or utilitarian grounds but are rather guided by normative standards from their culture and society. This study consequently aims to establish an ethical decision-making framework for moral dilemmas in driving situations that can then serve as a foundation for an obligatory ethical setting to be implemented in ADVs. Such a framework should prevent a system, able to save thousands of lives, not being used because of moral disagreements with the general population.

Studies including trolley dilemmas were traditionally carried out in the form of philosophical essays. This means that the material was presented to participants in the form of written scenario descriptions, sometimes with additional pictorial representations. This way of presenting the dilemma introduces issues, such as the disregard of important contextual and situational influences in moral decision-making (Skulmowski et al., 2014). New immersive technologies, such as Virtual Reality (VR), could help to remedy these insufficiencies. In this context, trolley dilemmas have recently experienced a revival in science (Navarrete et al., 2012; X. Pan et al., 2011; Patil et al., 2014; Skulmowski et al., 2014).

The immersion that VR environments provide serves to improve ecological validity while maintaining control over experimental variables (Madary & Metzinger, 2016). In the context of ADVs, VR can present scenarios that are more similar to real life decision-making in traffic and hence shed light on the moral actions of the participants rather than their conscious beliefs. Furthermore, many possible modifications of the trolley dilemma elicit open questions. For example, different characteristics of potential victims might influence the human decision process. Previous studies have shown that children were saved more often than adults, so the ages of potential victims might play a role in decision-making (Sütfeld et al., 2017). In the context of ADVs, there are also certain traffic-specific aspects worth considering. For instance, sidewalks provide a safe space for pedestrians in traffic which might lead to an internalized reluctance to drive on side-walks and could also influence the decision process in a modified trolley dilemma. Additionally, there are possible scenarios in which people can only save lives by sacrificing their own. Despite evidence from surveys that revealed a willingness to use self-sacrificing ADVs (Bonnefon et al., 2016), it is questionable whether people would indeed act this way in realistic settings. The present study specifically addresses these open questions and aims at a high ecological validity by using a VR setting.

In this experiment, five hypotheses were tested. First, based on previous research, it is postulated that people will, in general, act in favor of the quantitative greater good, trying to keep the number of persons to be hit to a minimum (Hypothesis 1). Yet, it can be speculated that the ages of potential victims matter in the sense that people might spare younger individuals at the expense of older ones (Hypothesis 2). In the traffic-specific context pedestrians on the sidewalk are expected to be protected, as they are not actively taking part in traffic. By staying on the sidewalk, people generally expect to be safe while implicitly giving consent to the finite risk of being injured when stepping into the street. Therefore, people are assumed to avoid hitting pedestrians on sidewalks as opposed to people standing on streets (Hypothesis 3). On the other hand, it is hypothesized that people prefer to protect children, even if they are standing on streets, as opposed to adults on sidewalks (Hypothesis 4). Finally, the last hypothesis states that people will not reject self-sacrifice completely but consider it when a high threshold of damage to others is reached (Hypothesis 5). To test these hypotheses, a driving simulation experiment with state-of-the-art VR technologies was implemented, following a study by (Sütfeld et al., 2017). The presented avatars were only male to avoid an effect of gender difference, as previous studies showed that male and female avatars are treated differently (Sütfeld et al., 2017). Participants were able to control cars as drivers and experienced various modified trolley dilemma situations, as specified in the "Methods" section.

5.2 | Human Decisions in Virtual Dilemma Situations

Results

Data from 189 participants and a total of 4000 trials, distributed into five modules according to the aforementioned hypotheses, was analysed. Below, the results for each module will be described separately.

Quantitative Greater Good

In the first module, it was tested whether people would act in favor of the quantitative greater good by saving more as opposed to fewer avatars. This module consisted of three trials. The environment for this module was a suburban setting, consisting of a two-lane road. Only standing adults were presented as avatars. In the suburban setting, parked cars occupied both sides of the two-lane street. In the one-versus-two and one-versus-six conditions, only 7 out of 189 participants targeted the higher number of avatars (Figure 5.1). In the one-versus-four condition, 12 participants targeted the group of four instead of the individual; thus, in all three conditions, the overwhelming majority of participants spared the larger number of avatars.

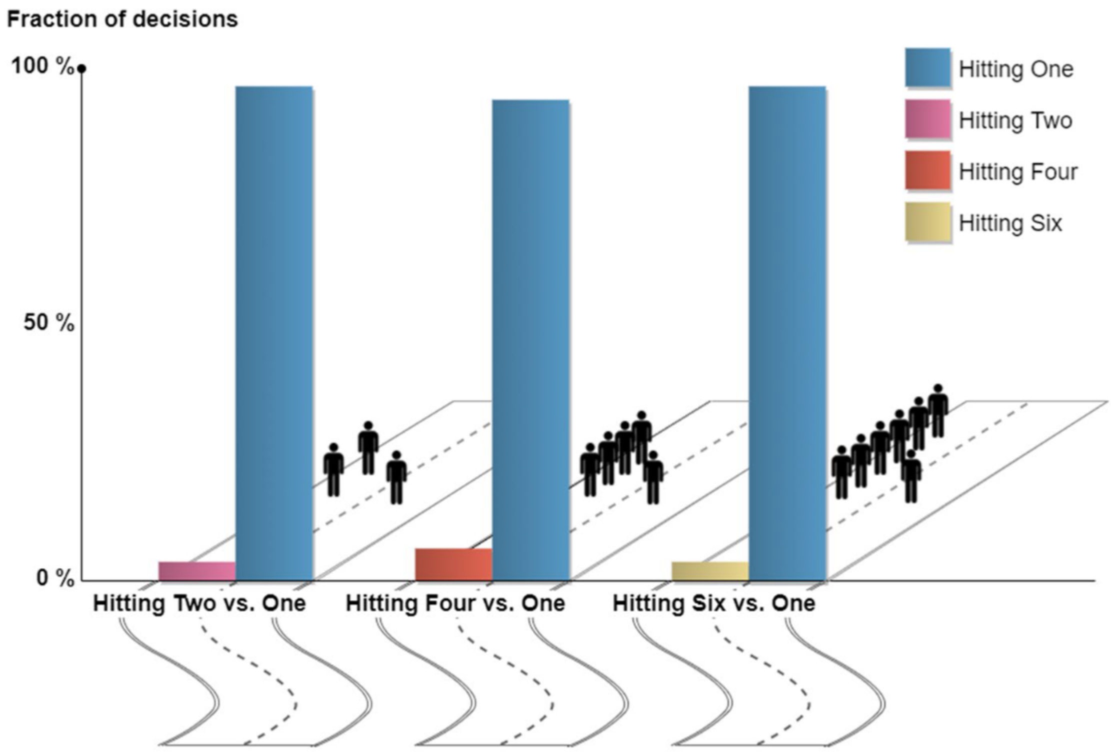


Figure 5.1: Decision distribution in the Quantitative Greater Good module

To investigate this difference between the conditions, a permutation test was used. It yielded no significant difference ($p > 0.05$). This shows that participants acted similarly throughout all three conditions. For each single condition, the number of participants targeting one avatar instead of the larger number is highly significant ($p < 0.01$). This data indicates that participants decided in favor of the quantitative greater good.

Age-Considering Greater Good

The second module tested the hypothesis that people would spare younger avatars at the expense of older ones. It was composed of six trials in a suburban setting. As avatars a child, a standing adult, a kneeling adult, and an old person were used. Each trial presented one of the following six combinations of avatars: one child versus one standing adult, one child versus one old person, one standing adult versus one old person, one kneeling adult versus one standing adult, one kneeling adult versus one old person, and one kneeling adult versus one child.

In the pairwise comparisons of children, adults, and the elderly, it was observed that younger avatars were spared at the expense of older avatars (Figure 5.2). The differences between children versus adults and elderly versus adults were highly significant in a permutation test ($p < 0.001$). The results demonstrate the inverse relation of the expected remaining lifespan of an avatar and the chance of getting hit. This decrease in value according to age was highly significant ($p < 0.01$). To investigate whether the difference emerged only through variation in avatar height, kneeling adults versus standing children and standing elderly were tested. The observed difference in the children versus kneeling adults' comparison was highly significant (Figure 5.2, 4th block, $p < 0.001$).

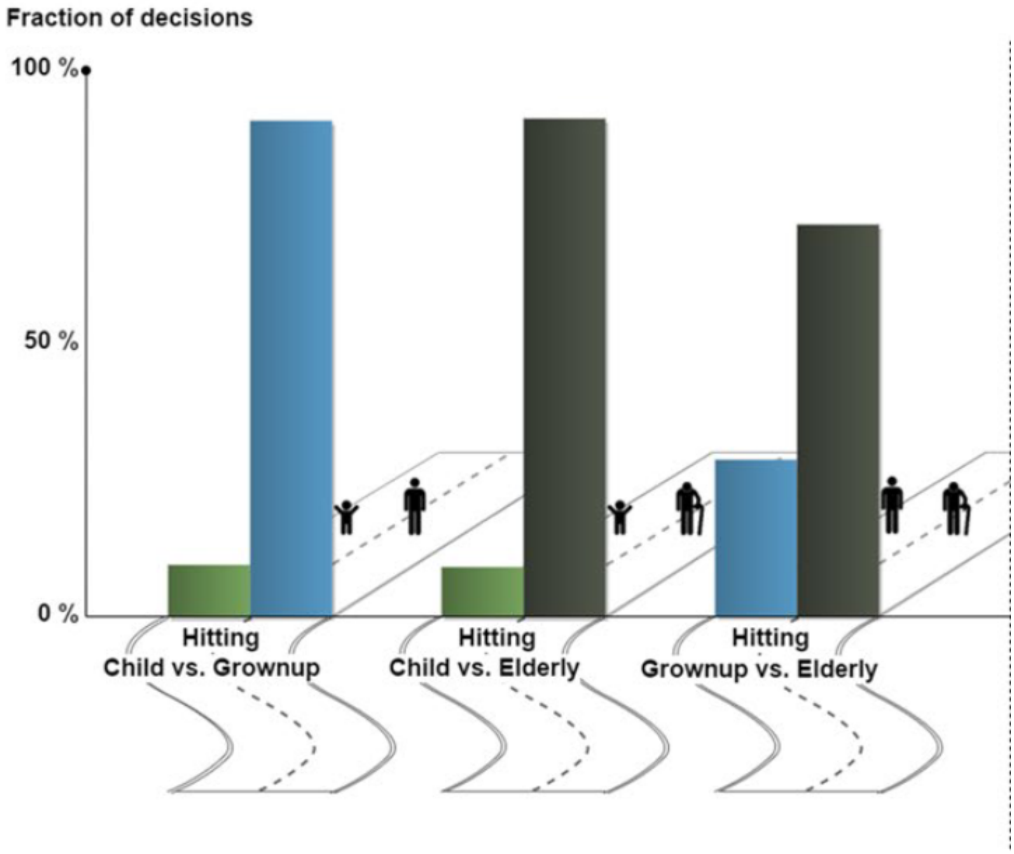


Figure 5.2: Decision distribution in the Age-Considering Greater Good module of purely age-considering decisions.

In the direct comparison of kneeling adults versus standing adults, the latter were hit more often ($p < 0.001$). A similar pattern emerged in the comparison between kneeling adults versus the elderly; thus, kneeling and standing moderated the participants' decisions to some degree. (Figure 5.3) However, these results confirm that participants spare younger avatars at the expense of older ones, irrespective of the avatars' heights.

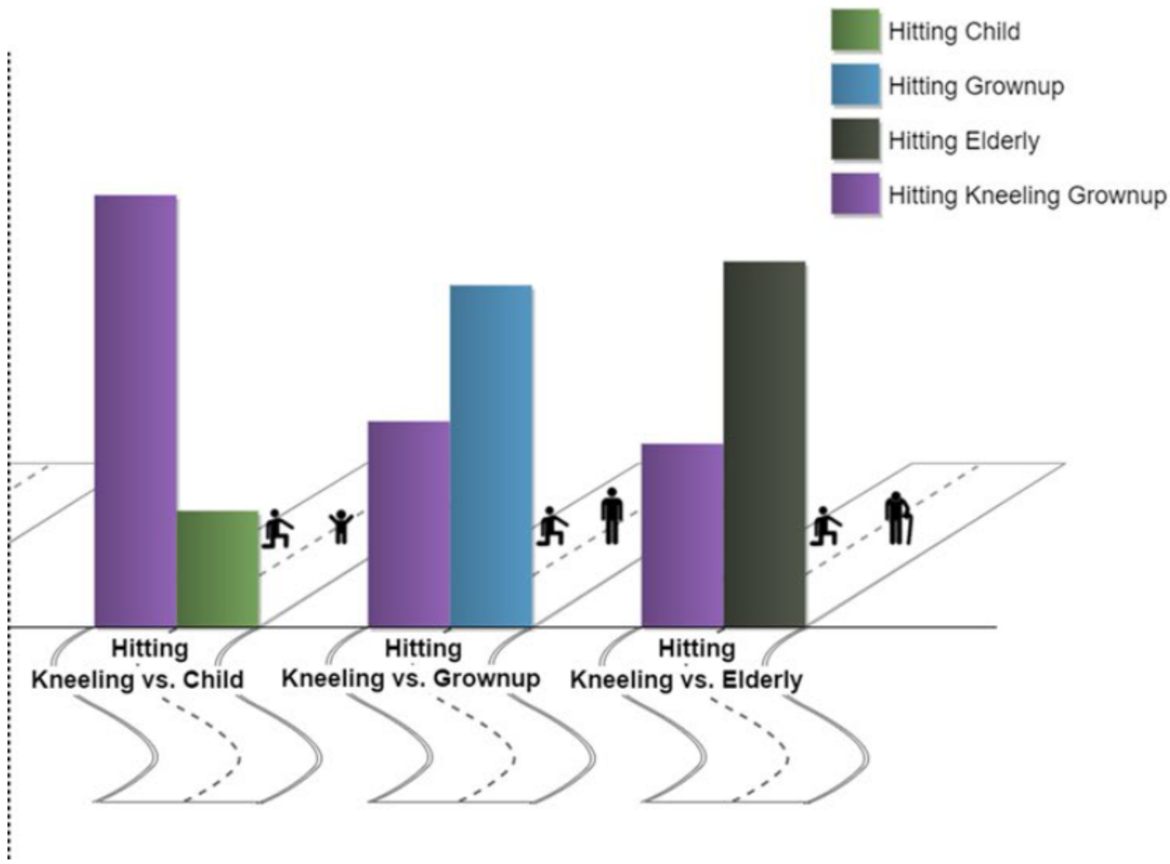


Figure 5.3: Decision distribution in the Age-Considering Greater Good module of decisions about object height.

The Influence of Context

The third module explored the influence of context. Specifically, the corresponding hypothesis states that avatars located on sidewalks would be spared more often than those in streets. Therefore, in direct analogy to the first module, a single adult avatar on the sidewalk was matched with two to six adult avatars in the street.This module consisted of six trials in a city setting that contained a one-way street with sidewalks on both sides. One of the sidewalks was blocked by parked cars while participants had the opportunity to drive on the other sidewalk to avoid avatars in the street.

Compared to the first module, it was expected that a larger difference in the number of avatars would be necessary to lead to a consistent sacrifice of the single avatar on the sidewalk. However, in general, this context did not seem to have a strong effect on decisions. The majority of participants still consistently spared the high-est number of avatars possible, regardless of the sidewalk context (Figure 5.4). It was investigated whether a switch point, defined by a critical imbalance of the number of avatars, could adequately describe the participants' decisions. That is, if the number of avatars to be hit in the street were larger than this threshold, participants would change from driving in the street to driving on the sidewalk to save a large enough group of avatars. The data showed that only 2.56% of trials would need to be changed for all participants to behave consistently according to a simple model with a single free parameter, the switch point.

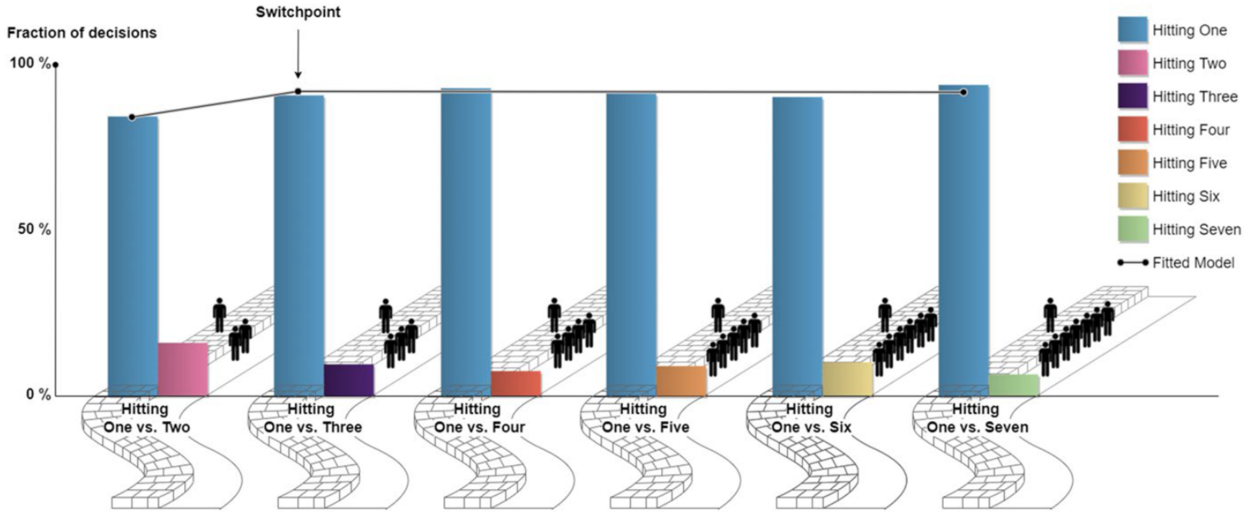


Figure 5.4: Decision distribution in the Influence of Context module. Depiction of the best-fitted model for these decisions.

For statistical evaluation, models describing different switch points were fitted to the data and compared the sums of squared residuals of the models to identify the model that best fits the data. Results showed that modeling the data with a switch point between the conditions with one-versus-two and one-versus-three avatars described the data best (Figure 5.4), with a sum of squared residuals of 34.0. This, in turn, indicates that participants choose to drive on the sidewalk to save a group of three or more avatars rather than saving only two. However, through-out all conditions, the number of participants who drove on the sidewalk to save more avatars was significantly higher than those trying to save the avatar on the sidewalk. In comparison to the Quantitative Greater Good module, only minor quantitative differences were found. This shows that the sidewalk altogether has a surprisingly small effect.

Interaction of Age and Context

n the Age-Considering Greater Good and the Influence of Context modules, the influence of age and context was investigated in isolation. The fourth module, was designed to find out whether there was also an interaction of age and context; hence, the city setting with the sidewalk, including child avatars, was used. There were three trials with the following combinations of avatars: two children in the street versus one adult on the sidewalk, one child in the street versus two adults on the sidewalk, and one child in the street versus one adult on the sidewalk.Results showed that the majority of participants again spared children as opposed to adults, despite the sidewalk context (Figure 5.5), as could be expected based on the findings from the previous modules.

In further analyses, two permutation tests were performed to check for differences in the target actions of participants regarding the number of avatars. The conditions with one child in the street and one or two adults on the sidewalk were significantly different from one another ($p < 0.001$). The same held for the comparison of the condition with one child and one adult versus the condition of two children and one adult ($p < 0.05$). The results were in accordance with the findings from all previous modules. Furthermore, the pattern of the results was compatible with the independent effects of the sidewalk and age.

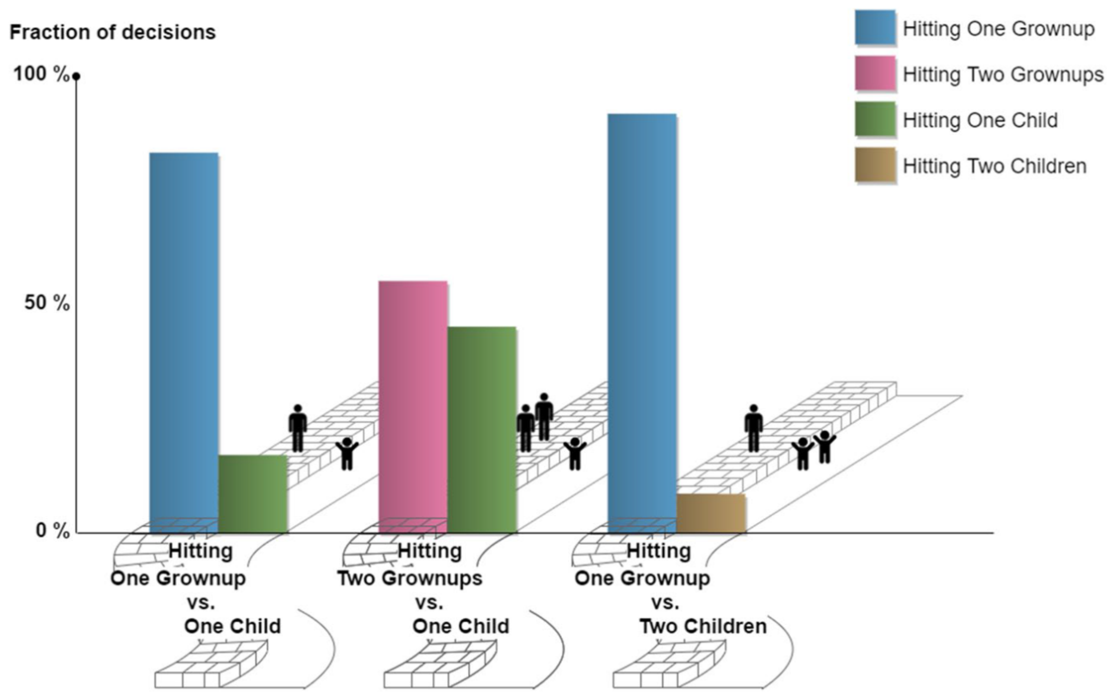


Figure 5.5: Decision distribution in the Interaction of Age and Context module

Self-Sacrifice

The fifth module investigated whether participants value their own life in the VR setup similarly to the value of other avatars. That is, they had the possibility to save avatars at the price of sacrificing their own avatar. In close analogy to the previous modules, participants' choices were investigated, depending on the number of avatars in the group opposing self-sacrifice. In comparison to the non-self-sacrifice condition in the first module, the switch point i.e., the number of avatars in the group necessary to induce consistent decisions, was hypothesized to increase.The Self-Sacrifice module contained six trials in a mountain setting, where a chasm was implemented on the right lane of the street with a construction sign in front of it. Presented on the left lane were varying numbers of standing adults, ranging from two to seven avatars. The design was created to imply that participants would commit self-sacrifice within the experimental paradigm by driving off the cliff when using the right lane.

Data analysis of this module followed the same procedure as in the Influence of Context module. It was postulated that a fixed threshold could describe the behavior of the participants. In case the number of avatars in the street was below the thresh-old, the group would be sacrificed. In contrast, when it was above the threshold, participants would choose self-sacrifice. The decisions of only 5.2% of trials were not consistent with such a simple model.

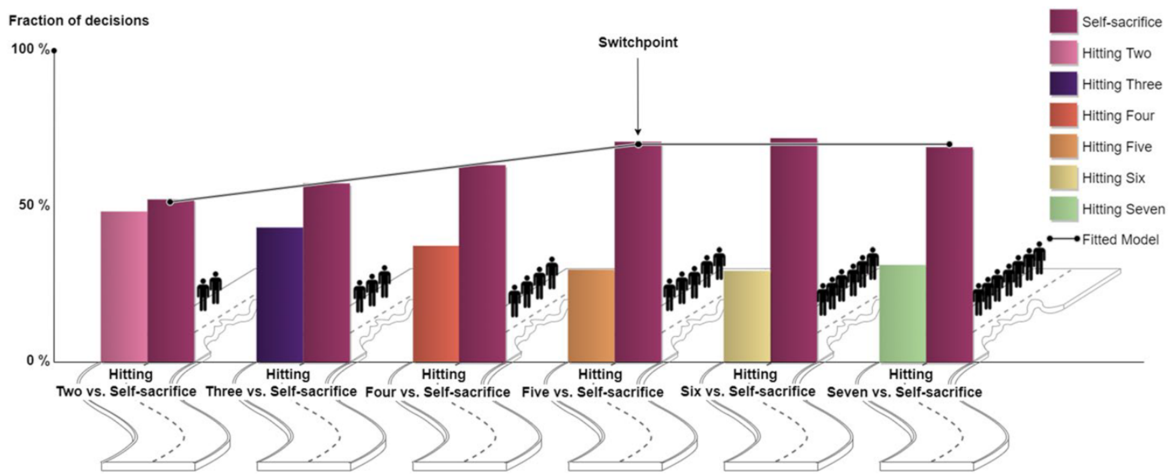


Figure 5.6: Decision distribution in the Self-Sacrifice module. Depiction of the best-fitted model for these decisions

To see if there was a general switching point at which the number of people committing self-sacrifice does not increase anymore, a linear regression was used to fit six models to the data. For these models, the sums of squared residuals were computed and compared. The model with a switch point between the conditions with four and five avatars best described the data (Figure 5.6). With a value of 3, the sum of squared errors of this model was much smaller than those of the other models. This indicates that people are consistently willing to sacrifice themselves in the case of being able to save a group of 5 or more avatars with this decision.

Methods

Participants

Two hundred sixteen unpaid subjects participated in the study. Participants were acquired from various venues throughout Osnabrück. Data from 27 participants had to be excluded from the analysis for various reasons: 15 participants did not complete the experiment due to nausea or disagreement with the experimental settings; 12 participants did not pass training (more than six trials). In the end, data from 189 participants served for analysis (62 female, 127 male). They were aged between 18 and 67 years with a mean of 24.32 years.

Stimuli and Design

In the VR environment participants were driving a car on a one-directional track with two lanes. The environmental surroundings varied between five settings. One suburban and two mountain settings consisted of a dual roadway, where the starting lane was randomized for each trial. The two city settings consisted of a one-way street and a drivable sidewalk; starting lane was always the street. The car was driving at a constant speed of 36 km/h, visible to the participants on the car display, and the track length ranged between 180 and 200 m to avoid habituation (Figure 5.7).

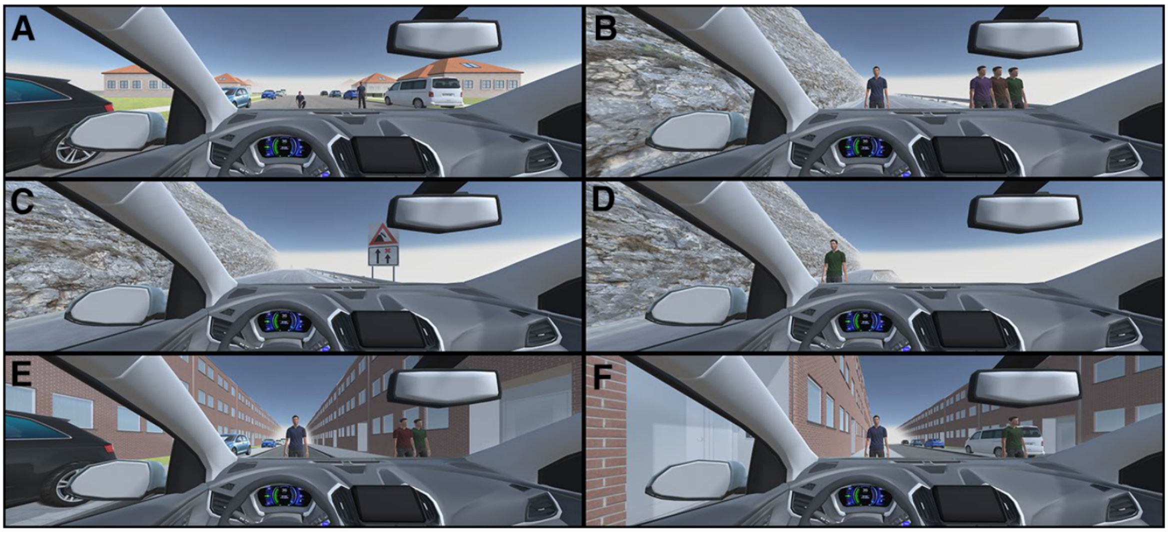


Figure 5.7: Screenshots of the VR environment in the different modules. a Age-Considering Greater Good module in the suburban setting. b Quantitative Greater Good module in the mountain setting 1. c Self-Sacrifice module showing the road sign warning of the oncoming chasm in the mountain setting 2. d Self-Sacrifice module in the mountain setting 2. e Age-Considering Greater Good module in the city setting 2. f Age-Considering Greater Good module in the city setting 1

At the end of each lane, distinct types of male avatars appeared in different combinations, forcing participants into a dilemma-situation. The lane side on which the two types of avatars appeared was randomized for each trial (excluding modules containing a sidewalk) and did therefore not correlate with the starting lane.

To decrease the visual range and thereby guarantee a constant decision-making time of four seconds, all settings included foggy weather. At the beginning of each trial, a beep indicated to the participants that they had control over the vehicle. The relatively low speed of the car was selected as a compromise to allow reasonable time for deliberation and to have the nature of the obstacle clearly visible. At the same time, it involves the danger that not all participants perceive a car crash as a threat to the life of the avatars. At 15 m from the avatars, another beep signaled that the control over the vehicle was withdrawn, as later inputs would have led to incomplete lane change maneuvers.

Five Hypotheses were tested using different experimental modules.The Quantitative Greater Good module consisted of three trials to test Hypothesis 1. The environment for this module included the suburban and mountain setting 1. There was always one standing adult avatar in one lane (randomized) as opposed to either two, four, or six in the other lane.

The Age-Considering Greater Good module aimed at testing Hypothesis 2. It was composed of six trials in the suburban setting. A child, an adult, and an old person were used as avatars. Additionally, a kneeling adult served to make sure that possible effects were not only due to the size of the stimuli, given that the kneeling adult was of the same height as the child. Each trial presented one of the following six combinations of avatars: One child versus one standing adult, one child versus one old person, one standing adult versus one old person, one kneeling adult versus one standing adult, one kneeling adult versus one old person, and one kneeling adult versus one child.

The Influence of Context module investigated Hypothesis 3 and consisted of six trials in the city setting with a one-way street plus a driveable sidewalk on the left or right side in the city setting 1 or 2, respectively. The setting was randomized for each trial. There was always one standing adult on the sidewalk and two to seven standing adults in the street, each combination occurring once for every participant.

The Interaction of Age and Context module testing Hypothesis 4 was based on the Influence of Context module also using the city settings, including children as avatars. There were three trials with the following combination of avatars: two children in the street versus one standing adult on the sidewalk, one child in the street versus two standing adults on the sidewalk, and one child in the street versus one standing adult on the sidewalk.

The Self-Sacrifice module investigating Hypothesis 5 contained six trials in the mountain setting 2. Here, a chasm was implemented in the right lane of the street with a construction sign in front of it. The chasm was at the edge of the mountain road and it was bottomless. A vehicle or person falling into the chasm would fall down the steep flank of the mountain. A varying number of standing adults, ranging from two to seven avatars, was presented in the left lane, while the chasm was in the right. It was created to imply that participants would commit self-sacrifice within the experimental paradigm by driving over the chasm when using the right lane.

Procedure

After giving written consent, participants were seated in front of a keyboard and equipped with the Oculus Rift DK2 in combination with Bose Noise-Cancelling Headphones. The VR experiment contained all instructions and consisted of three phases: training trials, experimental trials, and a questionnaire. Participants used the left or right arrow key to change the driving lane. The training phase contained three trials where participants had to avoid three pylons that appeared in one of the lanes alternately. If they hit a pylon, the trial had to be repeated. After successfully completing all three training trials, the various types of avatars were presented to the participants before the experimental trials started. Finally, a questionnaire was answered, which is beyond the scope of the present paper. The duration of the whole experiment was approximately 15–20 min.

Statistical Tests

For all analyses, only final decisions were taken into account.A permutation test was performed to investige the influences of the number of avatars on participants' decisions in the Quantitative Greater Good, Age-Considering Greater Good and Interaction of Age and Context modules for each condition individually. Additionally, a binomial test using pooled data proved significance in comparison to the null hypothesis of a random distribution of choices in the aforementioned modules.

For the Influence of Context and for the Self-Sacrifice module, fractions of trials that would need to be changed for each participant to show a consistent decision-behavior were calculated. A fraction of up to 5% could be explained by a natural error rate (Kuss et al., 2005). To test whether the data match the hypothesis of a general switch point, six models of different underlying switch points were fit-ted to the data and the performance of each was computed. Assuming that upon a certain switch point the number of participants committing self-sacrifice would not further increase, the mean between the conditions with an avatar number higher than a certain switch point was calculated. The model was assumed to pass through this mean in a plateau. For the conditions with avatar numbers smaller than the underlying switch point, a linear increase up to the calculated mean was expected. To test which model fits the data best, the sums of squared residuals were computed and compared.

5.3 | Developing Normative Ethics for ADVs

Discussion

When driving a car in VR, participants act in favor of the quantitative greater good. That is, their behavior consistently aims at sparing as many avatars as possible. This even applies to situations in which participants have to virtually sacrifice their own avatar to save others. Age and context modulate these behavioral patterns. Specifically, the probability to sacrifice an avatar and its expected remaining lifespan showed an inverse relation. Participants consistently saved younger avatars as opposed to older ones. Surprisingly, the context of sidewalk versus street only had a small influence. In conclusion, the results throughout all conditions support the hypothesis that people act in favor of the quantitative greater good, even in scenarios involving a sidewalk or self-sacrifice.

Concerning the results of the quantitative greater good module, it could be argued that the participants' decisions were guided by self-preservation, based on a higher expected probability for themselves to survive when hitting only one avatar com-pared to a group of avatars. However, results of the Self-Sacrifice module indicate that for most participants the influence of self-preservation is limited as already more than half of the participants are willing to commit self-sacrifice for only a group of two avatars. This led us to believe that the decision process was mainly guided by the aim to spare as many avatars as possible.

On the other hand, even though self-preservation seems to be limited, the results show that a considerable portion of participants value the life of their own avatar higher than the life of another avatar as reflected by the existence of individual switch points for avatar groups higher than two. The overall behavior in the self-sacrifice module could have also been influenced by the case that it was not self-evident that hitting one or more avatars would automatically lead to their death. Indeed, after the experiment some participants reported that the car was too slow to kill a human being. Before killing themselves by driving into a bottomless chasm in the self-sacrifice trials, the participants chose to drive into the group of avatars, risking injuries but not lethal damage. Although phrasing the trolley dilemma not in terms of life and death but in terms of health or injury is equivalent, it might lead to differences in decision-making. This should be taken into account when interpreting the results of the Self-Sacrifice module where injuring avatars is opposed to killing their own avatar.

Other complex processes can be assumed to underlie decisions in the age considering greater good module. In social sciences, the term disability-adjusted life years (DALY) is used (Murray, 1994). This is a complex measure that can be roughly understood as the number of years lost in a healthy life. Such a description naturally explains the inverse relation between age and the probability of being spared. Thus, the decision process might be better described not by simply counting the number of lives, but with a more complex measure, such as the DALY. The possible influence of measures like the DALY can be underlined by the control condition using kneeling adults instead of children, showing that the effect of sparing children in comparison to adults was not purely determined by the size of the avatar's visual appearance. Nevertheless, the aspect that a collision might not necessarily lead to the death of the victim has to be considered when comparing avatars of different ages, too. Besides age and expected lifespan, the likeliness to die in the case of a crash might have influenced the decision of participants to hit adults more often than children. This could also explain the preference to save kneeling adults in comparison to standing adults due to the increased risk of fatal injury when hit by a car in a kneeling position. Hence, it is possible that participants were pondering complex decision processes, including severity of injury and risk of death. These complex decision processes were, contrary to expectations, not substantially affected by a sidewalk, given that participants did not seem to be reluctant to drive on the sidewalk to spare a number of avatars higher than two.

In general, though, decision processes could have been influenced by a lack of incentives. Although VR provides a more realistic experience than surveys or pictorial descriptions, it can still, especially regarding self-sacrifice, be argued that participants in contrast to real-world situations did not have to fear any consequence for their actions. However, it is not possible to implement an experimental feature that represents self-sacrifice or killing other people in a realistic and ethically acceptable way. It is, therefore, questionable whether people would actually be willing to sacrifice themselves in order to act in favor of the quantitative greater good when it comes to such a dilemma situation in traffic or whether their behavior only emerged out of social desirability. This leads to another important aspect, namely that whatever is socially desirable might considerably differ between societies. It is to be noted that since the majority of participants in the current study are Germans and the moral behavior of people with different socio-economic and political backgrounds can significantly vary, the results cannot easily be transferred to other societies.

Moreover, the behavior of participants could have been affected by limitations of graphical display and therefore immersion. This contrasts with some participants dropping out of the experiment because they did not feel comfortable with hitting or even killing virtual avatars. The latter observation does not support a lack of realism or immersion. However, there seem to be many individual differences in play. In this regard, it cannot be ruled out for sure that some participants, especially young ones, were not as committed to the study as expected but were mainly interested in the new VR technology offering a game-like experience. Thus, the average degree of immersion was high, but individual variations should be taken into consideration in future research addressing these problems.

In the field of implementing autonomous driving behavior, empirical knowledge is relatively sparse and ethical approaches are widely debated. Usable ADVs as well as advanced simulation techniques, like 3D VR, are relatively new. Consequently, empirical studies rely heavily on questionnaires directing issues straight at potential customers. The behavior of ADVs and their control algorithms will be judged by the standards and ethics of the societies in which they operate. This again emphasizes the crucial role of acceptance, because self-driving cars need moral algorithms capable of expressing three aspects: being consistent, not causing public outrage, and not discouraging potential buyers (Bonnefon et al., 2015). For example, the Head of Active Safety of Mercedes Benz, Christoph von Hugo, stated that Mercedes would only build ADVs that would consequently save the driver of the vehicle in hope that this would make the car more attractive to buyers (Morris, 2016). But the morality of such automated vehicles should be questioned as there is a gap between what people state how they want an ADV to behave and what they would actually buy (Bonnefon et al., 2016). It is debatable whether ADVs should be programmed based on economic reasons instead of human behavior or ethical arguments.

As mentioned before, traffic is a complex interaction between many road users. The choice to always save one's the own life in a critical situation affects the future response of other road users. There may be a lot of people who would choose a self-sacrificing ADV, but if the majority uses a self-preserving car, this will change. In cases like these, the result could be dramatic for society since the chance of being killed in traffic would rise (Gogoll & Müller, 2017). In the following lines, an argument for an obligatory ethics setting in ADVs will be developed. It will also be explain, why a modified trolley dilemma, like it is used in this study, is suitable for a foundation of such a regulated ethics setting.

The above mentioned standpoint of Mercedes Benz represents a moral egoist standpoint. Such a standpoint is plausible if the driver cannot be sure how another car will react in the case of a crash. If the passenger is not disposed to act in favor of the greater good, why take the chance of being killed by a stranger who might act selfishly? This would lead to a prisoner's dilemma like situation. Each road user could choose between self-sacrifice as an analogy for cooperation in the prisoner's dilemma, or self-preservation as an analogy for defection. To maximize the good for the society, it would be adequate to choose a possible self-sacrifice. This would be the lowest toll for society, like in the prisoner's dilemma the lowest combined sentences. This could be an ADV acting in a utilitarian way, as it sacrifices the passengers for the greater good no matter how another road user would act in a crash situation. But if one road user has the possibility to stay alive while the other road user sacrifices her- or himself, the result would likely be that most people prefer the ethics setting of self-preservation to prevent being exploited by people driving, for example, a selfishly programmed vehicle. Hence, even if an individual would like his or her ADV to act in a utilitarian way in traffic, the outcome for society would be worse due to the clash of different ethical settings. For example, an ADV with a strong passenger preservation setting might push a school bus full of children into an abyss to ensure the safety of its own passenger. Such a tragedy could be avoided by the same ethical setting for all ADVs, like a utilitarian one. Therefore, one could argue for an obligatory ethics setting, a form of governmental intervention as a common standard for the behavior of ADVs in crash situations. As Gogoll and Müller, 2017 showed, a summarized maxim for such a standard could be to minimize harm for all the people involved. Moreover, this argument also does not allow mixed traffic as an interim solution since this comes close to the situation of personal ethics settings which would lead to the aforementioned prisoner's dilemma, too.

A potential solution to this dilemma resides in a contractarian stance. The present findings of participants acting in a utilitarian way would, however, probably not last long if transferred to a real traffic scenario with a variety of road users and an according variety of intentions in possible crash scenarios. In this case, the observed behavior of participants corroborates the results of the thought experiment by Gogoll and Müller, 2017. In the present study, participants' behavior could be described by utilitarianism more than self-preservation. Yet, not only saving the passenger of the vehicle at all times but also programming a vehicle to behave in utilitarian ways, even in dilemma situations not including self-sacrifice, is contradicted by German law. The first para-graph of the first article of the German constitution states that the human dignity is inviolable, which implies that all humans are equal and contradicts any evaluation of human life. This forbids the saving of one human at the expense of others, as well as any quantifying perspective on human lives. Even if possible solutions for the implementation of completely autonomous cars are contradicted by the law, it is important to discuss these issues in order to advance towards a legal solution in the future.

According to Morris, 2016, the statement of Mercedes Benz was later revised due to the mentioned legal issues. Nevertheless, ADVs not only have to embody the laws but also the ethical principles of the society they operate in (Gerdes & Thornton, 2015).

Regarding the issue of self-sacrifice, it is arguable whether the trolley dilemma is an adequate setting for the moral problems of ADVs. As Gogoll and Müller, 2017 state, the trolley dilemma is missing strategic interaction and iteration, meaning that the participants' actions alone determine the result of the dilemma situation. The participant's decision is independent of other human actions. The decision does not take responses of other road users into account. Simply spoken: The situation in a real-life trolley-like scenario is not only determined by a personal ethics setting, but by all involved road users and their ethics settings. The goal of this study was, how-ever, to develop a derived ethics setting which most people agree on. The only way to be able to see what people agree on, even if it is only due to social acceptance, is to isolate the strategic intentions and iterations from the possible scenario. In these cases, the participant's decision is independent of the reaction of different road users as well as of social acceptability since only the participant was able to see the result of her or his action.The present findings suggest that the gap between a set of mandatory ethical rules, which solve the prisoner's dilemma and the behavior of the participants is not as big as one would expect.

Despite the case that various studies were conducted under the assumption that people would like ADVs to behave similarly to humans (Goodall, 2014; Malle et al., 2015) concerns could be raised about the uniform behavior of ADVs. Sikkenk and Terken, 2015 found that many factors dramatically influence human behavior in traffic, e.g. weather conditions and the driving style of other traffic participants. Variation does not only occur in driving behavior but also in judging decisions of humans in contrast to those of machines (Malle et al., 2015). This was also shown by J. Li et al., 2016 who examined the differences in responsibility between humans and machines in cases of inevitable fatal crashes. Participants had to judge the decision of either a human driver or an autonomous car in a dilemma. In contrast to human drivers, where utilitarian decisions were most favorable, participants expected ADVs to behave in a utilitarian manner under all circumstances.

Different studies point out that the general population seems to favor utilitarian decisions (Bonnefon et al., 2015; J. Li et al., 2016; Malle et al., 2015). This applies even to cases in which drivers have to sacrifice themselves for the greater good (Sachdeva et al., 2015). Such behavior can be understood as an act of maximizing utility (Thomson, 1985). Therefore, it is mostly referred to as utilitarian reasoning and decision-making. Ab initio, utilitarian decisions offer themselves as a quantitative treatment and appear to be suitable for ADVs. However, the problem of how to implement ethics in machines, especially in dilemma situations, remains. A recent project at the Bristol Robotic Laboratories (Winfield et al., 2014) tried to implement Asimov's three laws of robotics[2] to show that robots can be ethical, as well as safe. In this experiment, they defined a puck-like robot as a human which moved towards a hole in a table. Another robot, the one with the implemented robotic laws, had the task to save the robot that moved towards the hole. But the experiment showed that there was no such thing as a simple rule, like the first Asimov law of robotics, to save a human life in dilemma situations. In these situations, the robot showed inconsistent behavior when it should rescue two robots moving towards the hole. Some-times it was able to rescue one, sometimes even both, but in the worst scenario none was saved. To fix that, more rules have to be applied to the initial code. But who is the one to be saved first? Could there be a rule to prioritize one life over another? A simple solution is not in sight but seems to be a crucial aspect in promoting ADVs to potential customers and allowing them to be an integral part of society. Ethical dilemmas do not necessarily have absolute answers, but they do have significant ethical implications for users. Ethicists serving as experts for ethical evaluations of robots are key to solving the question of how ADVs should behave (Millar, 2016). Despite several unresolved issues, there are many ethical arguments for fully autonomous cars. Not only is it likely that they improve mobility for elderly and disabled people, but also reduce crashes and annual fatalities in traffic, ease congestion, allow more work-related activities while driving, and therefore a productivity gain, improve fuel economy, reduce parking issues, and offer transportation to those unable to drive. For example, the US economic benefits could reach around 25 billion dollars per year with only a 10% market penetration. Including high penetration rates, this raises the annual benefit up to 430 billion dollars, which makes ADVs a technology for a better future (Fagnant & Kockelman, 2015). The number of avoided fatalities is a sufficient reason to promote ADVs. Therefore, the idea of McBride, 2016 for partial automation only is decidedly rejected here. Instead, further research addressing open questions should be encouraged. These range from technical issues to ethical and psychological problems, as well as legal aspects, such as responsibility and policy issues.

[2]The three laws of robotics (Asimov, 2013) were created as a part of a science fiction novel by Isaac Asimov as a concrete beginning of possible ethical settings for robots. They are human centered, and easily applicable to ADVs as well. 1. A robot may not injure a human being or, through inaction, allow a human being to come to harm. 2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law. 3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

In summary, the results show that participants consistently behave in utilitarian ways in various dilemma situations. Their decisions were only slightly modulated by context, such as a sidewalk. Furthermore, the effect of age might be subsumed in an utilitarian decision process as well. Even in conditions involving a self-sacrifice, participants' decisions were compatible with a utilitarian strategy. The study describes driver decisions in possible traffic dilemma situations more accurately than a mere survey. In contrast to previous surveys, it is shown that people are inclined to act in a utilitarian way. The majority of participants acted in a way that their behavior would minimize harm to all road users. This maxim could be derived as a mandatory ethics setting in all future ADVs, because a personal ethics setting could result in a prisoner's dilemma situation and more fatalities in traffic. The option of implementing such a setting is more likely to be accepted if people indeed act in the way discovered as opposed to the way they describe their actions. The study provides a basis for an algorithm implementing morals regarding ADVs. It describes how human car drivers would behave in these conditions and what is therefore seen as adequate behavior in general traffic situations.

Acknowledgements

The authors would like to thank all study project members: Aalia Nosheen, Max Räuker, Juhee Jang, Simeon Kraev, Carmen Meixner, Lasse T. Bergmann and Larissa Schlicht. This study is complemented by a philosophical study with a broader scope (Larissa Schlicht, Carmen Meixner, Lasse T. Bergmann). The work in this paper was supported by the European Union through the H2020-FETPROACT-2014, SEP-210141273, ID: 641321 socializing sensorimotor contingencies (soc-SMCs), PK.

Author Contributions

This study was planned and conducted in an interdisciplinary study project supervised by Prof. Dr. Peter König, Prof. Dr. Gordon Pipa, and Prof. Dr. Achim Stephan. Maximilian Alexander Wächter, Anja Faulhaber, and Silja Timm shaped the experimental design to a large degree. Leon René Sütfeld had a leading role in the implementation of the VR study design in Unity. Anke Dittmer and Felix Blind contributed to VR implementation. Anke Dittmer, Felix Blind, Silja Timm, and Maximilian Alexander Wächter contributed to the data acquisition, analysis, and writing process. Anja Faulhaber contributed to the data acquisition and the writing process.

Financial Interest

This publication presents part of the results of the study project “Moral decisions in the interaction of humans and a car driving assistant”. Such study projects are an obligatory component of the master’s degree in cognitive science at the University of Osnabrück. It was supervised by Prof. Dr. Peter König, Prof. Dr. Gordon Pipa, and Prof. Dr. Achim Stephan. Funders had no role in the study’s design, data collection and analysis, the decision to publish, or the preparation of the manuscript.

5.4 | Chapter Summary

In this chapter, we have seen that the test subjects behaved primarily utilitarian. The decisions made may be influenced by the context of the traffic situation, but they remain subsumable under utilitarianism. Similarly, the age of the other road users had a minor influence on the decisions made by the subjects. The majority of participants just acted so that their behavior minimized harm to all road users or maximized years of life.

Therefore, we argue that the benefit of the larger group could serve as a general guideline for future vehicles. This maxim may become a mandatory ethics requirement in all future ADVs. Personal ethics settings implemented to reflect the users' values could lead to more traffic fatalities. Personal ethics may lead to a prisoner's dilemma situation, where a suboptimal outcome is chosen to prevent exploitation by other road users. However, the presented study shows that there was a considerable consistency over participants toward utilitarian decisions (Faulhaber et al., 2019). Therefore we argue, that utilitarian decisions are considered appropriate behavior in general traffic situations (Bergmann et al., 2018). Similarly, further studies could show that people act utilitarian themselves and demand this of other drivers when they evaluate the situation as observers. What is exciting about this is that people especially demand utilitarian behavior from autonomous vehicles (Kallioinen et al., 2019). Thus, utilitarianism not only seems to be a practical maxim for performing socially acceptable actions in dilemma situations, but it also objectively helps minimize harm in society. In summary, a utilitarian vehicle reflects the values of the majority of society and contributes to society's good by minimizing casualties. Thus, as an easy-to-implement action guide, utilitarianism meets the requirements as standard programming for self-driving vehicles (Faulhaber et al., 2019).

There has been much development in jurisprudence on this topic in the past few years (Greger, 2018). In particular, Germany drafted a new traffic law that is particularly capable of providing information to make autonomous vehicles a reality on our roads (Bundesministerium für Digitales und Verkehr, 2017). Even though the utilitarian approach offers an implementable solution that minimized fatalities, there are unresolved issues. On the one hand, because the implementation of an mandatory utilitarian approach means discrimination against certain individuals. If a vehicle selects individual road users in an extreme situation, and in case of doubt, the elderly road users as collision targets, this represents discrimination which is not compatible with our constitutional Law (§ 1 paragraph 1 GG). Of course, it is possible to discard any personal information like gender and age, but still the car would actively target individuals compared to a group. This itself is a problematic implementation since quantification in weighing human life for the protection of another is not consistent with our constitution (Bundesministerium für Digitales und Verkehr, 2017; Bundesministeriums der Justiz und für Verbraucherschutz, 2013). Thus, utilitarianism is hardly possible basis for a normative ethic of self-driving vehicles.

Another even more severe problem is that the ethics Council considers any fixed programming for self-driving vehicles incongruent with the German constitution. The reason is that any decision that is recorded in code is intentional. For example, in the case of targeting in road traffic, this could be punished as intentional homicide (Bundesministerium für Digitales und Verkehr, 2017; Hevelke & Nida-Rümelin, 2015a). Thus, it is not advised to implement generalized decisions in a dilemma scenario.

However, what does this mean for the goal of this thesis? The course of the first chapter described the underlying technology of self-driving vehicles. The taxonomy of self-driving vehicles according to the SAE was discussed as well. It was shown which activities are necessary for the driving task and in which classification the portions of the driving task on humans and the machine are divided. In the connection of these topics, it could be shown that the driving task turns out to be more difficult for machines than intuitively often assumed since the vehicle does not only have to make calculations about its own behavior. Instead, it continuously compares its own movement with other human road users, whose behavior is difficult to predict.

Similarly, we have seen that the current legal basis does not allow the vehicle to drive without a driver (Greger, 2018). Thus, from a legal point of view, vehicles of automation levels 4 can only be implemented in public traffic, if there is a human supervisor involved. Therefore, it appears that the current level 3 will remain in place until a legal and technical solution for fully autonomous driving vehicles is found. Until that time, the development will likely remain in level 3, in which the driver can take over tasks that are not relevant to driving but has to respond quickly to a takeover request. However, the human remains in charge of all the vehicle's actions. As we have seen in this chapter, the machine is not allowed to make vital decisions independently. It has also been shown that humans in level 3 vehicles have a longer reaction time and may not intervene quickly and in time in traffic events. Therefore, we see the irony of automation since the technology decreases driving safety while increasing automation that is thought to prevent accidents. Because of these remaining issues, the driving automation might face a winter until the full automation is achieved.

How can a standstill in development be prevented in the coming years? One solution that will be proposed here is based on improving driver response by reading desired actions from driver behavior. In summary, with the studies shown in this thesis, we could demonstrate that it is possible to test self-driving vehicles and human-machine interaction with them in virtual reality and create realistic environments that enable valid research of the HMI in the context of ADVs. We have seen that the way a vehicle communicates can also increase the trust of passengers toward the ADV. This can be shown via a self-report of a simplified TAM version and the analyses of head movements of a large data set of participants. Likewise, we are still evaluating the extent to which a driver can react faster and more appropriately in a takeover using a human-centered HMI in an augmented reality approach in virtual reality. At least there is supporting literature that shows similar results (Jarosch, Gold, et al., 2019; Wintersberger et al., 2020).

Accordingly, in the next section, we will develop a proposal for a human-centric and semi-autonomous HMI. Here, existing automation technology is used to detect the human driver's intentions before the driver can act. Subsequently, the machine, which now knows about the driver's intention, intervenes in the driver's actions to make the decision and action of the human driver more precise. This action is thought to stabilize driving maneuvers if the human reacts too violently or too little due to lack of time. We will also discuss whether it is still possible to simultaneously solve the ethical component and the legal issues with such an approach. Ideally, we will be able to build a framework in which it is possible to build such an experiment in the coming years to show that we can at least partially circumvent the current weaknesses and hurdles on the way to automation.

6

Towards Multimodal Human-Machine Cooperation

Nezami, F. N.†, Wächter, M. A.†, Maleki, N., Spaniol, P., Kühne, L. M., Haas, A., Pingel, J. M., Tiemann, L., Nienhaus, F., Keller, L., König, S. U, König, P. & Pipa, G. (2020, May 25). ***From Interaction to Cooperation: a new approach for human-machine interaction research for closing the out-of-the-loop unfamiliarity. (Preprint) PsyArXiv; 2020. doi: 10.31234/osf.io/7jg3c.***

† shared first authorship

6.1 | From Interaction to Cooperation

In the previous chapters, various aspects of the interaction between humans and ADVs were explained. These chapters identified problems that may be a potential barrier to the further development of self-driving vehicles - ranging from the very foundation of this technology to the human-machine interaction. Yet, a closer look also revealed opportunities to transform potential problems into new technical possibilities in each chapter. The following part aims to summarize the previous chapters and create a new research idea based on the gathered knowledge. The suggested idea is called SensARa: a machine-learning-based HMI for human-machine cooperation thought to solve issues of the OOTLU in the case of a fast and precise takeover, as well as legal and ethical issues of automated decision making for level 3 automation.

As seen in the first chapter, self-driving vehicles are thought to be the next step toward large-scale processing of vast and complex data by a machine in the middle of our society (Dajsuren & van den Brand, 2019; Ryan, 2020). Yet, technical, legal, social, and ethical challenges might slow down or even prevent the development of vehicles automation higher than level 3. As described in the introduction, the automated vehicle will be confronted with uncertainties in sensors readings or the untrained environment. These uncertainties may lead to false assumptions of the system and are therefore a possible source for accidents (Fagnant & Kockelman, 2015). When it comes to a fatal decision, it is not even clear on which basis the ADV has decided since the vehicle's algorithm is most likely opaque. This makes the question of guilt and responsibility irresolvable. Furthermore, while a system may be designed for accident avoidance, there is still a small chance of inevitable accidents (Faulhaber et al., 2019). Therefore, current systems cannot replace a responsible driver or supervisor at assessing and evaluating a critical traffic situation. (Bundesministerium für Digitales und Verkehr, 2017; Hevelke & Nida-Rümelin, 2015b).

Nevertheless, it was shown that ADVs are desirable since they most likely decimate fatal accidents in individual traffic (Campbell et al., 2010). Also, they allow for more environmental-friendly mobility as fewer cars are needed, and a more inclusive individual traffic, since no driving action, and therefore no driving license, is required anymore (Chehri & Mouftah, 2019; Cusumano, 2020; Hars, 2016). Since acceptance in the general public is still low, it is essential to raise awareness of opportunities and risks in order to create understanding and acceptance in this technology (C. Lee et al., 2019; Waytz et al., 2014). One approach discussed in this thesis is to build a human-machine interaction providing information regarding the vehicle's actions. This can be done in any modality, as long as it enables functional understanding of the system, but effectiveness differs. (Endsley et al., 2003; Koo et al., 2015; Koo et al., 2016). Communication about the current state and decision of the car increase acceptance and trust. However, the right information at the right time is also thought to bring the driver back into the control loop in case the driver is distracted. At the end of the introduction, virtual reality was introduced to investigate a possible human-machine interaction in a safe and realistic yet controlled environment. Here, testing self-driving functionalities and research on the human-machine interaction in critical traffic events are possible, while the test subjects' reactions are monitored precisely.

In the second chapter, the presented project LoopAR intends to examine the human-machine interaction in the context of level 3 self-driving vehicles, where a human driver can engage in NDRTs but is still responsible for all driving actions of the automation features (Dix et al., 2021; Greger, 2018). The study aims to examine if a multimodal HMI can decrease takeover time while increasing the precision of the human response by guiding human attention. This study included a modular and open access unity toolbox with sufficient realism and size to resemble realistic driving environments conform to traffic regulations. LoopAR is thought to foster VR research and serve as mobile and cost-efficient simulator replacement. We could show that relevant research can be achieved in mobile and cost-efficient simulator replacement using customer-grade VR equipment. Additionally, we presented the code required and the assets used in this environment that enable interested users and researchers to easily and quickly adapt virtual studies in the context of vehicle automation.

To test whether in-vehicle verbal feedforward and feedback cues given by the car can increase trust and acceptance in self-driving vehicles, we created a virtual driving experience with more than 26000 participants, as seen in Chapters 3 and 4. After the experimental trial, we asked subjects to rate the ease of use, perceived usefulness, intention to use such a car, and trust in the automation. We were able to show that verbal information given by the car increases trust and perceived usefulness, which is supported by prior research (Baltes et al., 1990; Bengler et al., 2020; Endsley et al., 2003; Koo et al., 2015; Wintersberger et al., 2020). Nonetheless, we were also able to show that trust does not seem to be the deciding factor for acceptance since we saw a negative effect on the intention to use such a vehicle and the perceived ease of use in the data. Also, we found a substantial difference in the questionnaire scores between men and women and a decrease in acceptance with increasing age.

However, we saw that there is no generalizable solution for the problem of how a self-driving car should communicate with people. Instead, we found that people can be divided into three groups with different adversarial needs. The first group - the doubtful - is defined by low scores in ease of use and intention to use. It can be assumed that the skepticism towards ADVs is not purely due to the technology itself. Rather, we suspect that it has to do with technical competence. As seen in the study from Chapter 4, the reason for this is the high scores for trust and perceived usefulness, but the low scores for ease of use.The second group is defined by uncertain users, with low intention scores in the AVAS condition but higher scores in the silent automation without verbal cues. Here we assume that this group is open to automation but finds the additional communication annoying and prefers the conventional radio to the self-explanatory vehicle. The final group is the group of overestimators with the highest scores overall conditions. Previous research also identified this group of ADV users, which is defined by unrealistic expectations about the ADVs' capabilities. The overestimation of the vehicles capabilities might be a substantial factor in fatal accidents (Gillmore & Tenhundfeld, 2020; Penmetsa et al., 2021; Wintersberger et al., 2020).

While analyzing the recorded head movements of the participants, we were able to show that subjects exhibited larger but slower movements when they were more accepting and faster but narrower movements when they were less accepting. In addition to the explicit attitudes from the questionnaires, we could also show that the head movements during the experimental trials can be used as a valid measurement to examine implicit attitudes. From this finding, the idea arose to create an adaptive HMI for level 3 automation vehicles that can alter its communication style based on the behavior of a human passenger. As the head movements of the human driver can be associated with the need for information about the current traffic situation, the car would assess these movements and adjust the information given accordingly so that the human driver is supported optimally in a fast takeover situation. In an automated driving situation, this could enhance trust and acceptance since the car could quickly react to the needs of the out-of-the-loop passenger while simultaneously enabling fast and precise reactions of the human supervisor in a takeover maneuver.

Nevertheless, the development of the tested self-driving functionalities still faces legal and ethical obstacles. Therefore, we conducted another virtual driving experiment, where we tested if a forced-choice decision in a dilemma situation can be summed up under an ethical theory. We found that the tested subjects decided in a utilitarian way. However, participants were influenced by the context of the traffic situation. Similarly, the age of the other road users had a minor influence on the decisions made by the subjects. The majority of participants just acted so that their behavior minimized harm to all road users or maximized years of life. Therefore, we conclude that minimizing damage, and in dilemma situations also minimizing the loss of lives can be a possible implementation if the decision is not based on personal characteristics such as gender or age (Bundesministerium für Digitales und Verkehr, 2017). Therefore, utilitarianism not only seems to be a practical maxim for performing socially acceptable actions in dilemma situations, but it also objectively helps minimize harm in society (Faulhaber et al., 2019).

The following and last section will be used to develop a human-machine interaction for level 3 automation based on the previous chapters. This interaction is intended to turn the human and the machine into cooperation partners through mutual knowledge about each other. This should solve the problems raised and enable further development of the technology.

Introduction

The question of how to design a user-friendly interaction for complex automated systems like automated cars has been the subject of scientific discussion for decades (Bengler et al., 2020; Endsley & Kiris, 1995; Norman, 1990). The main reason for this question is that highly automated vehicles of level 3 and above will change the demands on the driver's cognitive system in a radical way (S. Li et al., 2019). The role of humans as physically active decision-makers in vehicles will be replaced gradually by automated systems (Lindgren et al., 2020), such as the Lane Departure Warning System (H. J. Kim & Yang, 2015) or Forward Collision Warning System (Haber, 1978). As seen throughout the thesis, with the growing efficiency of technical systems with a human fallback, human-centered interaction is becoming increasingly important. The safety and reliability of an automated system, where humans and machines share the DDT, cannot be achieved by optimizing technical components alone (Schreurs & Steuwer, 2015). Instead, the reliability of automated systems is primarily determined by the quality of interaction between humans and machines as an intertwined process (Altendorf et al., 2017; Marberger et al., 2018). As previous research suggests, in-vehicle communication about the vehicle's state and the environment can lead to a better understanding of the actions of the vehicle (Koo et al., 2016; Wintersberger et al., 2020). Additionally, it is suggested that augmented reality visualizations and warnings could be a promising technology to foster trust and acceptance of algorithmic decisions (Nezami et al., 2021; Wintersberger et al., 2020). Up to now, this is used to give the human a representation of the vehicle. The ideal cooperative system should include representations of both interaction partners to ensure cooperation. This means that humans should understand and predict the vehicle's behavior, while the car can interpret and predict human behavior.

Throughout this thesis, we saw that while automation can increase safety and comfort, its implementation is not without risk. One source of failures is a lack of communication, when the human does not know about the vehicle's actions and is therefore not able to assess a traffic situation. Consequently, the responsible driver acts inadequately. Still, not only the driver has too little knowledge about the vehicle. It will be argued that the vehicle also has too little information about the driver to enable satisfactory decision making. Since humans and machines have only limited representations of each other (Waschl et al., 2019), we aim to develop design recommendations to reduce the risks of misunderstandings in the human-machine interaction on both sides. This applies in particular to situations where humans have the task of taking over system control in the event of sensor failures or malfunctions; the OOTLU (Endsley & Kiris, 1995). Thus, investigating the fluent and fast control switch in the takeover is of crucial importance for the safety of autonomous cars (Melcher et al., 2015).

As seen in the previous chapters, the driver in level 3 automation must take over immediately after a request, even when not engaged in driving-related activities. If the takeover is unexpected, there is most likely an orientation phase in which the driver tries to assess the current traffic situation (Merat et al., 2014). Unfortunately, the driver's reaction is often too slow in accident situations, as there is only a small time frame of fewer than four seconds before an impact occurs (Green, 2000; Summala, 2000). Even in the case of fast reactions within said brief interval, studies have shown merely hectic responses by human drivers, which of course did neither improve the situation nor the outcome (Melcher et al., 2015). As an alternative to improve and shorten the takeover time and support reasonable actions, we suggest supplying targeted information while discounting potential distractors to ease the orientation phase and facilitate the driver's reaction via an adaptive HUD.

Combining the chapter findings of this thesis, we can formulate a proposal for a human-centered design. This is thought to increase the understanding between humans and machines. Therefore, this HMI can keep the human driver in the control loop and represents a possible solution to the ethical and legal problems. In addition, it should thus be possible to enable a kind of trust calibration through the additional information and therefore reduce a possible overestimation while increasing trust in doubtful users (Stephan, 2019). Finally, we will propose to enhance human actions by automation, creating a humanized automation as human-machine cooperation rather than a fully autonomous system without human drivers. In a critical situation, the machine can read the human intention and enhance the human reaction that would have otherwise been uncontrolled because the human driver is out-of-the-loop.

The European Union has formulated regulations on how the driving task shall be safely handed over from the assistance systems to the driver, including the capacity for the system to come to a stop if the driver does not reply appropriately. The regulations include requirements for the HMI to prevent misunderstandings or misuse by the driver. Specifically, the regulation requests a driver availability recognition system to detect if the driver is present in a driving position, check for the safety belt, and driver availability for a takeover. This includes eye blinking, eye closure as well as head and body movements (Official Journal of the European Union, 2021). In principle, this means close monitoring of the driver is already possible today and has even been approved by the EU for automated functions. Up to now, these regulations of automated driving allow only under specific conditions, for example, when the driver is available, and the vehicle is driving on roads where pedestrians and cyclists are prohibited (Official Journal of the European Union, 2021; SAE International, 2014). Additionally, a physical separation should divide traffic from opposite directions. If any of the conditions of the driver availability are not met, the system shall immediately initiate a takeover request. Yet, the regulation itself states that in an ongoing emergency, the deactivation can be delayed until the imminent collision risk disappears (Official Journal of the European Union, 2021). These time-urgent acquisitions are an open backdoor to moral dilemmas. Especially in delayed takeover situations, it can be assumed that the human driver can no longer act appropriately in the short time available. So it remains to be said that driver monitoring is already possible today and even mandatory for automated vehicles of levels 2 and 3 in certain situations. It is also possible for a vehicle to resolve a critical situation without the human driver if the latter is not prepared to take over. Even if it remains open, how exactly a possible reaction looks like (Official Journal of the European Union, 2021). Nevertheless, it can be stated that the premises necessary for our proposal regarding the close monitoring of the driver during a takeover are fulfilled.

Nevertheless, adaptive HMIs should not be limited to additional information to the driver. In addition to recording the driver's movements, feedback from the vehicle to the driver about the human actions is also essential to enable human-machine cooperation. The background is the pragmatic turn in cognitive science (Engel et al., 2013). In this sense, human cognition is defined as an activity that is only possible with interaction with the external world (Wilson, 2002). A central concept is sensorimotor coupling, according to which motor actions are controlled by matching expected and perceived sensory stimuli. This development matches the rapid progress of machine learning techniques in recent years. As mentioned in the chapter about the perception of self-driving cars, deep neural networks are a standard in computer vision and a critical method for the perception of self-driving vehicles (Georgevici & Terblanche, 2019) and complex systems prediction (J.-F. Chen et al., 2014), where artificial systems often outperform human performance (Waldrop, 2019). Therefore, ADV's ability to perceive and predict other road users and the passenger will be combined with the premises of the embodied cognition to merge them in a human-centered adaptive HMI that provides information as well as motor assistance to enable cooperation.

The proposed design is based on the LoopAR toolkit in chapter 2, where eye tracking data serve as an AR feedforward visual stimulus directly into VR. The HMI should be enhanced with machine learning algorithms to direct the driver toward the desired action via specific motor feedback. Accordingly, the car should warn the driver with additional sensory input, but it should also enhance the driver's motor response by interpreting intention and executing the desired action. This project aims to bind together humans and machines in becoming true cooperation partners in joint action. Thus, the vehicle reacts to the actual and predicted driving movements and the NDRT-related behavior behind the wheel. In addition, the system uses information from the environment to ensure safe driving. We hope to shift the paradigm of future research from only visual-auditory-haptic cues to full sensorimotor integration assistance.

Materials and Methods

The experiment needs to be divided into two phases to use machine learning as active assistance. It includes an initial experiment, where participants' data is recorded and later used to form a machine belief about the environment. This initial recording is already done. Within the last year, we gathered data from more than 200 participants, in which eye tracking data, head position and rotation, steering wheel angle, velocity, pedal angles, and the test subjects' outcomes in critical traffic situations are recorded. As depicted in Figure 6.1 LoopAR covered the initial Setup of calibration and driving training and stored eye tracking and driving data during a drive in four different virtual environments with a large number of different road users. The Motor AI drives a large amount of the experimental trial. During the automated sections, the vehicle maneuvers the vehicle safely. In addition, we implemented critical traffic events to test whether an HMI could bring the distracted

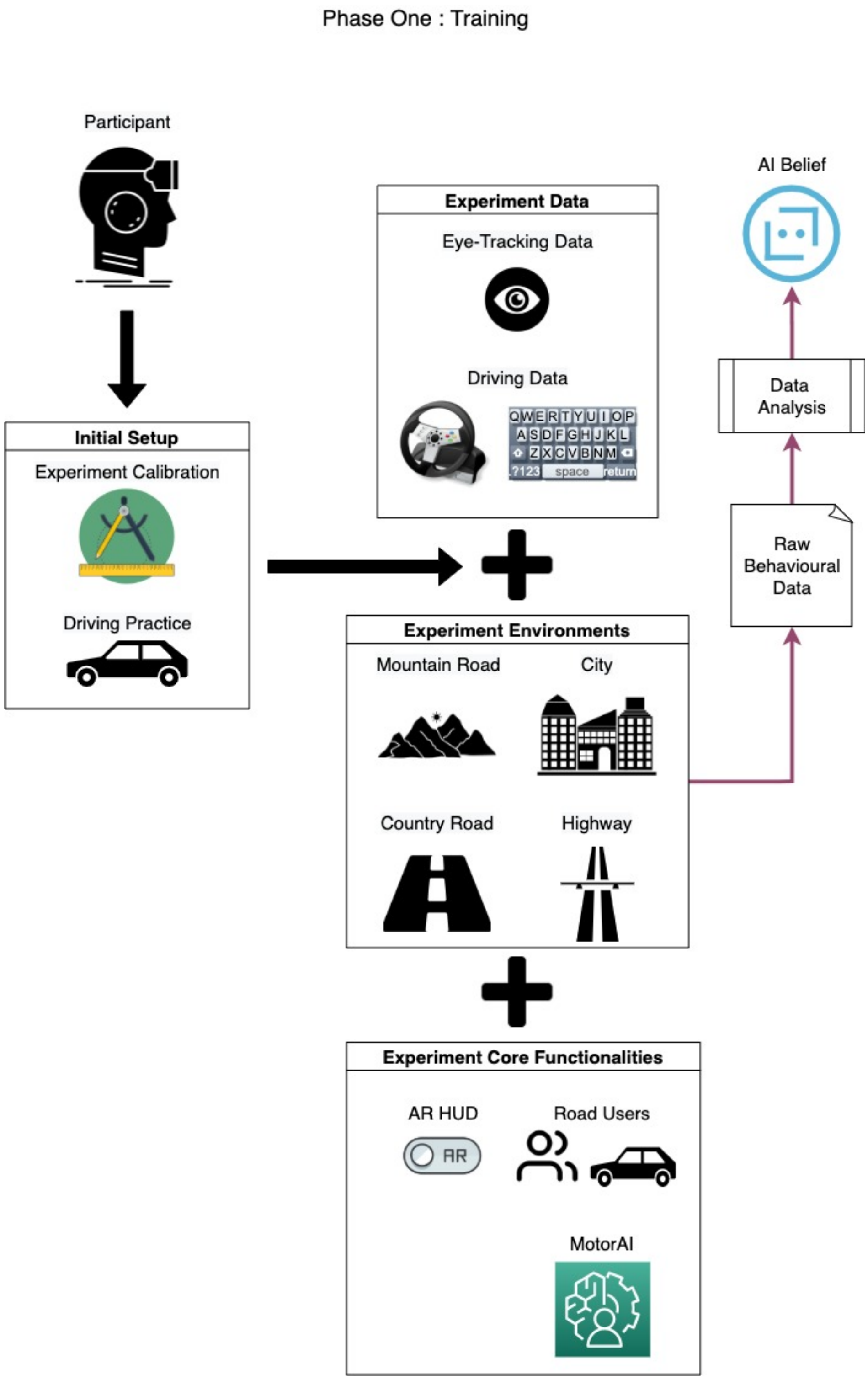


Figure 6.1: Overview of the training phase for collecting participants' data in four different environments

driver back into the loop. During 12 critical traffic events in the experimental trial, the car handed over control to the human driver 2,5 seconds prior to the critical situations. The driving data within the traffic events were stored as well as the post-study questionnaire. Depending on the condition, the experiment included an augmented reality HUD that highlights potential collision partners in critical traffic situations. In the second phase, the car AI will be trained with the behavioral data from phase one. This data is used to shape an AI Belief and a new set of participants will go through almost the same experiment as phase one. Gathered data of the test subjects in critical traffic situations will be used to shape the car's AI belief about desired actions and outcomes (Figure6.2). Participant's input from phase one will be fed to the AI Belief, which will assist the driver by correcting inappropriate responses. This module is called SenseAI (Figure 6.2), and will operate with the help of the already implemented MotorAI which controls the car outside of the critical traffic events.

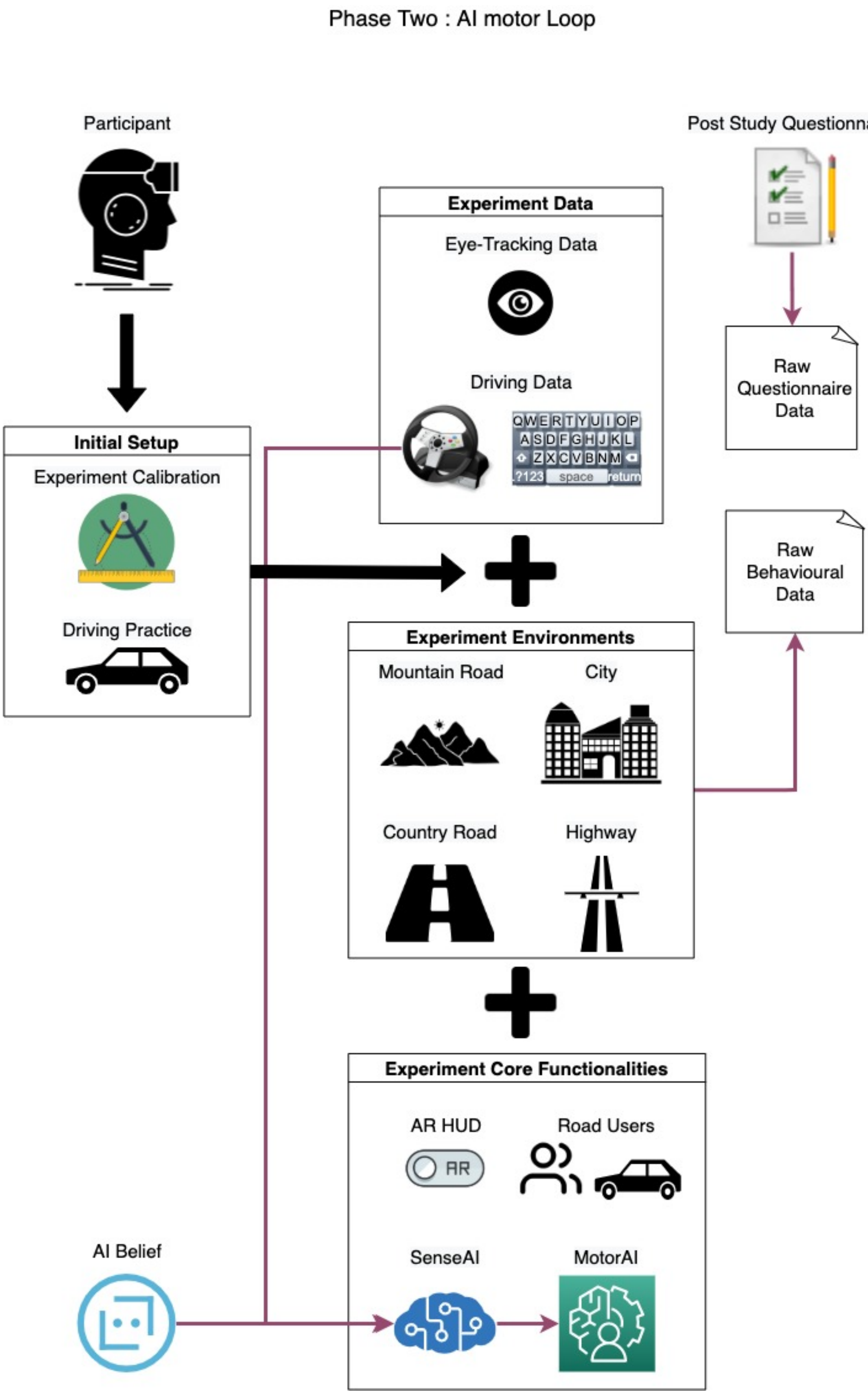


Figure 6.2: Overview of the SensARa output of the SensAI module

However, SenseAI will not be trained on subject data alone. Trained convolutional neural networks are thought to enhance it. CNNs already exist for the classification and localization of environmental objects in self-driving cars (Ouyang et al., 2020; Rao et al., 2019; M. Yang et al., 2019). These networks extract objects from the environment. With the identified objects, it is feasible to extract possible actions from the currently available environment of the vehicle. These actions are called vehicle affordances and relate the environment's features, including other road users and their actions, to the vehicle's possible actions in that situation. Up to this point, this describes the current state of the art of automated features. A first step further towards a transparent HMI is to make the objects identified by the vehicle visible to the human driver. To do so, critical objects will be highlighted in the augmented reality HUD used in the LooAR project (Nezami et al., 2021; Wintersberger et al., 2020).

As mentioned before, SensARa will also include a human adaptivity: this is an analysis of the driver's behavior in the form of eye, head, and hand movements in real-time. With the help of support vector machines, it is possible to predict intended actions up to two seconds in advance for the given context (Keshava et al., 2020; Keshava et al., 2021). This is thought to enable intention recognition, as the car is able to interpret the behavior of its passengers before the onset of an action. These functionalities will be coupled with the audio-visual feedforward and feedback of the LoopAR toolkit. Therefore, it is possible to select the affordances according to the vehicle affordances, the human affordances and the human intentions. Thus, it is primarily possible to enable a quick reaction in takeover scenarios since the attention can be focused on specific objects in the environment, which are crucial for accident avoidance.

The underlying idea stems from the simple use of Bayesian inference. It presupposes that the system has prior knowledge of the probability distribution of certain parameters. In the example presented here, the parameters include possible actions (e.g. braking or steering) of the driver or the autonomous car. In Bayes' theorem, these actions are defined as the prior. We use the test subject's inputs of the first phase of the experiment to define the prior. In the second phase, these inputs will be used to update the behavior of the AI car. The AI belief module (Figure 6.2) calculates the probability (post-hoc) of a particular driver action p(action) in a critical event. It uses the previous test subject inputs as an update factor (prior) and constantly multiplies it with the drivers' input (likelihood) during a critical situation. Since the input is multiplied every instance where the participant provides input, the probability of the action which is in line with the participant's input, becomes more and more pronounced during the critical situation. Although, this rather simple system is blind to certain objects and only allows testing in already defined traffic situations without moving road users. Thus, the system does not know whether the potential collision object is a pedestrian or a tree, but only estimates the likelihood of a collision. Therefore, for a first experimental trial, such a system could be used, but should be replaced, or supplemented, by the more complex methods. CNNs are able to identify pedestrians as such and also estimate possible movements of other road users (Ouyang et al., 2020). Thus, the system gains accuracy and is able to perform finer coordination, such as the protection of vulnerable road users.

Still, it remains open how this could possibly be a solution for the legal problems. Should SensARa intervene in traffic events based on the affordances and the human intention alone, it will remain a machine decision without human action. An independent decision, without an action from the driver, must not exist. Therefore, driver intention should be used in a further step not to replace human actions but rather to improve them. In the next step, SensARa should be able to optimize the drivers' performance with the system's knowledge of the user's state and intention, rather than replacing it (see Fig. 6.3). With the additional knowledge of the sensimotor coupling, it is possible to guide the movement execution of the drivers intended and executed action, enabling faster and precise reactions. This is proposed, since the human perception is thought to be closely bound to the action (Clark, 2006; Noë, 2009). The system uses environmental and user data in real-time to capture the intention and possibilities for action to improve the actual input data. The steering and pedal input is then adjusted toward the desired action, considering possible positive outcomes. This is realized with motor feedback cues in the steering wheel and pedals. This way, the driver's actions are not overruled but enhanced.

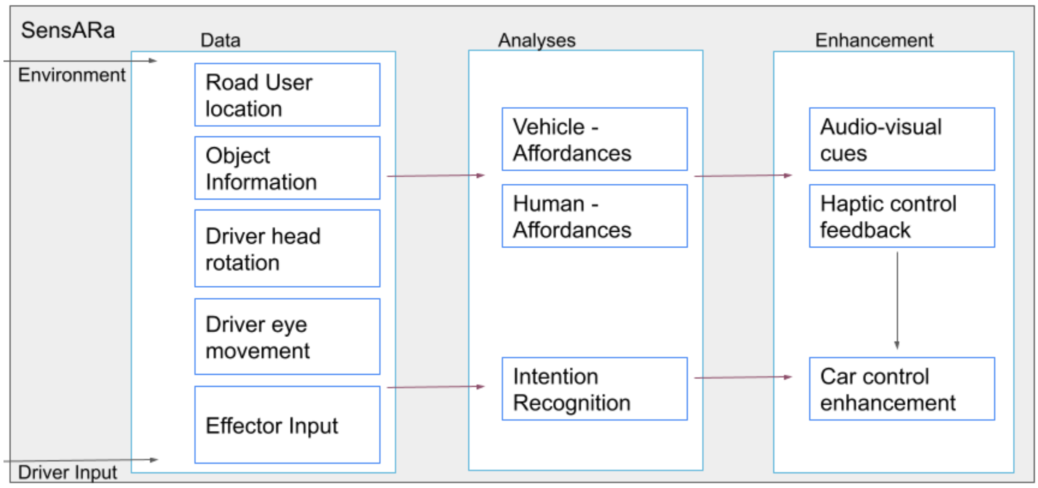


Figure 6.3: Component overview over the SensARa software architecture

Now, the question is how exactly this system can be embedded in a possible experiment. We hypothesize that subjects will not be able to respond immediately and appropriately in a time-sensitive event. We assume, for example, that a steering movement will be performed too excessively or too inert when the participant is under acute time pressure. We do not imply that the drivers' actions are random. Instead, we suggest inappropriate behavior of either too large or too little movements.

Together with intention recognition, these inappropriate movements allow the AI to smoothen or enhance the action toward more plausible and desired actions. The AI-belief will use the input from the vehicle affordance, the human affordance, and the intention recognition to identify and act according to the human will. The SenseAI module then uses the result to adjust the driver's reaction towards a probable outcome. The SenseAI does not take action on its own but instead infers the intended action based on its data-driven prior distribution on possible actions and current participant's input as well as the machines' predictions of users' input. Behavioral output together with a questionnaire will then be used to investigate the effectiveness of this cooperation, checking the outcomes of critical traffic events against the intended outcomes of the participants (Figure 6.2). Now that we have explained the theoretical setup, we will present a possible experimental setup. Afterward, the possible strengths and weaknesses of the experiment will be discussed.

Apparatus and Setup

For the technical apparatus, we suggest a computer equipped with an Intel(R) Xenon® E5-1607 v4, equivalent or better, 32 GB RAM with a 64-bit operating system Windows 10 Professional. An Nvidia GeForce RTX 2080Ti, served as a graphics processing unit to ensure a stable rate of about 90 frames per second. Similarily, we suggest the the HTC Pro Eye with a sampling rate of 90HZ, including the HTC base stations 2.0, as a VR setup and a customer-grade gaming steering wheel as the input device. The setup used and presented here has already been tested intensively in other VR studies and has proven to be very efficient in three large studies with over 12 months of continuous operation in the MS-Wissenschaft, LoopAR, and EEDA projects.

Experimental Procedure

As mentioned before, the proposed experimental design for SensARa is based on the LoopAR environments subject in chapter 2 of this thesis. Therefore, SensARa uses almost the identical procedure. A training trial scene is presented after the initial eye-tracking and seat calibration. Here the test subjects are introduced to the virtual car and the HMI. Participants are asked to take over manual control and maneuver through an obstacle course to complete the training. After completing the training scene, the experimental scene starts. At the beginning of each scenario, the car drives fully autonomous within the legal speed limits, which are being displayed in the head-up display that serves as HMI later in the experiment. In correspondence to the car speed, the scenes vary in length from 1,200m in the city scene to 3,600m on the highway. Each scene is loaded separately to enable a recalibration of the eye-tracker if the accuracy decreases due to participants' head movements. Every scene entails three critical traffic situations in which the test subjects are required to take over control of the car. After completing the trials in all four scenes, participants are asked to fill out a questionnaire on demographic data, the level of immersion, their previous VR experience, and their previous hours of gaming, which concludes the experiment. In addition, as a lesson from the incomplete datasets of the MS Wissenschaft, we use the AVAM questionnaire.

Stimuli and Critical Traffic Events

Stimuli

As stimuli, we use the information within the augmented reality head-up display provided by the autonomous vehicle (Figure 6.4). To decrease the takeover time of the driver, we give visual and auditory warning signals - just like in the previous LoopAR project. The stimuli appear when a takeover request starts. Therefore, the takeover request starts with a beep followed by a verbal warning (Veitengruber et al., 1977) accompanied by a red warning triangle (ISO 3864) on the windscreen. Following the initial warning, the visual warning sign blinks for two seconds to cue the driver's attention (Endsley et al., 2003). While the warning sign is flashing, the lettering informs the driver of the takeover request in a written form. The warning sign disappears after two seconds to give the driver an unobstructed view of the situation, only marking possible obstacles (Figure 6.4 bottom left). Finally, the car will grant control to the human driver 0.5 seconds later. The takeover request is finalized, and a small icon in the lower-left corner indicates that the driver is in complete control.

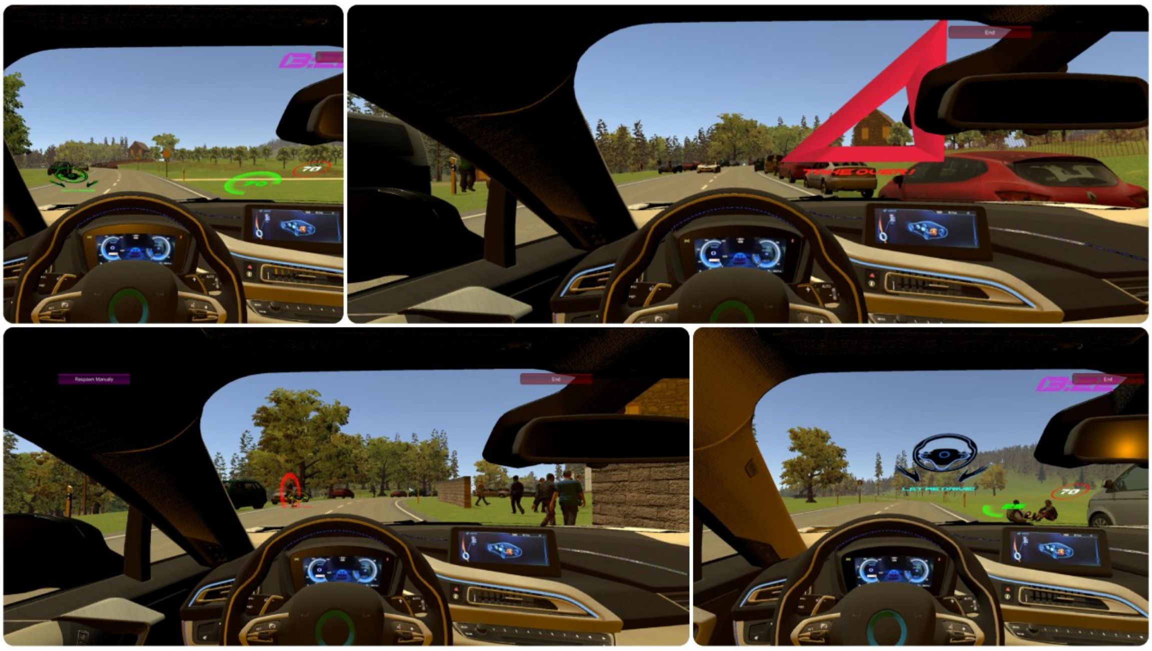


Figure 6.4: Panel of AR HUD in the proposed HMI

Critical Traffic Events

To assess the test subject's performance in terms of take-over time and reaction precision, we investigate driver reactions in a total of 12 critical traffic events. A take-over request is demanded when a car detects a critical situation that an artificial agent cannot resolve correctly without human intervention. This could be a deer jumping in front of the car, or a broken car blocking the road. These events depict situations in which the car cannot solve the situation within the legal boundaries. One example of a legal boundary is that an autonomous car is not allowed to cross a continuous line on the road, which might be necessary if an obstacle blocks the car's lane. In this case, the control should be handed to the human driver. In both of our experimental phases, we will, therefore, compare the outcomes of the critical traffic events and the quality and time of the driver's response. Once the critical traffic event is overcome, the car suggests a reinitialization of the autonomous driving mode in the form of a visible pictogram accompanied by text in the windscreen (Figure 6.4 bottom right). Another beep sound (with lower frequency) indicates successful transmission of control back to the car. The event zone prefab is designed to be easily customizable for different critical traffic event scenarios and environments.

Environment

The LoopAR environment is based on real geographical information of the city of Baulmes in the Swiss Alps. The region is selected due to its variety in terrain, including a small village, a country road, a mountain pass, and a region suitable for adding a highway section, totaling around 25km² of environment and an 11km drive on different roads. We decided to include these environments because urban areas, rural areas, and highways are the three main areas of car transportation. These areas demand different driving skills from an automated driving vehicle and a human driver. To make the region accessible in Unity, we used the collaborative project OpenStreetMap (OSM) (OpenStreetMap Foundation, 2004). We used the open-source 3D software Blender (Blender Foundation, 1995) to import the region and to extract the data. This provided information about vegetation, different street types, bodies of water, and buildings on the terrain, which we could use as references to build a close-to-reality, natural scenery. There is an entire network of streets that connects the scenes. Many of the assets used are from Unity's asset store and the 3D platforms Sketchfab and Turbosquid. Each of the four scenes can be customized anytime. It is therefore possible to change the number, size, and shape of all objects in each scene. Like previous work of the WestDrive project, LoopAR was developed for a special investigation but served as a foundation for various VR research.

Platform

SensARa is a further development of LoopAR, which is built in the Unity Editor. This software is a game engine platform based on C# by Unity Technologies, which supports 2D, 3D, AR & VR applications. The Unity editor and the Unity Hub run on Windows and Mac, and Linux (Ubuntu and CentOS), and built applications can be run on nearly all commercially usable platforms and devices. Unity also provides a large variety of available application programming interfaces and is compatible with numerous VR and AR devices (Juliani et al., 2020). Additionally, Unity enables the developers to access the Unity asset store, which is a large library of various assets such as 3D Objects and scripts that can be used to create versatile projects.

Traffic Behaviour

All cars used are based on the Unity standard wheel collider of the Unity3D physics engine. In the Car Core Module, user input is translated towards the motor control of the participant's car. The input consists of the motor torque, brake torque, and steering applied to the wheels. In addition to that, the Car Core Module can be accessed similarly by the AI Control Module. This allows a seamless transition from autonomous to manual driving. Furthermore, a gear system simulates varying torque on the wheels, increasing stability in steep road sections. Separate AI behavior modules handle the traffic behavior of the cars. The AI can follow predefined paths. Speed limit triggers inside the scene manipulate the AI's aimed speed, which is handling the input propagated to the Car Core Module (Figure 6.2). Another behavior allows the AI cars to keep their distance from each other. Currently not implemented is an obstacle avoidance behavior. The goal is to create an easily configurable and interchangeable traffic AI for multiple study designs. We maximize the realism of the car physics and the traffic simulation with these measures while ensuring easy adjustments. Paths followed by AI Cars and walking pedestrians were defined by mathematical bezier curve paths (J.-w. Choi et al., 2010), which were realized by a creation tool from previous Westdrive Projects.

Critical Traffic Events

Several trigger components realize the critical traffic events that are the core component of the SensARa trials. These independent triggers are activated when the participant's car enters the start trigger. The event zone is restricted within triggers. These triggers get activated if the car crashes into them, which is considered as participant's failure and leads to a black screen followed by respawning the car at a point after the event and giving back the control to the car. Furthermore, hitting the marked objects by the HUD also leads to respawning. The successful outcome of the scenario is when the participant enters the end trigger without crashing. The triggers are all visible in editor mode and invisible to the participant.

Behavioural Data

Gathered experiment data consists of the participant's input source and eye-tracking data. The eye-tracking component comprises calibration, validation, and online gaze ray-casting, which can record necessary gaze data during the experiment. The component is built for the Tobii HTC Vive Pro Eye device but is intended to keep the VR component interchangeable. Currently, it is intended as a simple connector to tap into SRanipal and the Tobii XR SDK (Figure 6.2). The eye calibration is performed with the built-in Tobii eye calibration tool. The validation is set in the corresponding validation scene, which provides a simple scenario with a fixation cross. Validation fails if the validation error angles exceed an error angle of 1,5° or the head was moved by 2" from the fixation cross. The information about the eye orientation, position, and collider hits will be stored with a calculated gaze ray of both eyes during the experiment. Currently, it is set to receive information about any object inside these rays to prevent the loss of viable information by objects covering each other. Scene-relevant information like the number of failed critical traffic events will be added shortly. All data is stored using generic data structures, serialized into JavaScript Object Notation (JSON), and saved with a unique participant ID at the end of each experimental block. The generic data structure used in the project ensures flexibility, as different data types can be added or removed from the serialization component. A calibration manager-script stores information about the seat calibration, eye validation error, and relevant test-drive results of every participant. This approach guarantees the highest compatibility with different analysis platforms such as R or Python.

6.2 | A Call for Future VR Research

As seen, SensARa enhances the currently existing LoopAR experiment with machine learning methods to enable the active assistance and augmentation of human behavior in highly automated vehicles. Therefore, it gives the test subject optimal sensory input and assists the participant with the motor behavior to ensure a fast and precise response of the human driver in case of a takeover request. The machine is guided by the possible actions in the directly perceived environment and the human driver's actions that are considered probable. With the knowledge of possible actions and the driver's intentions, the system can augment the driver's subsequent actions.

However, the exciting part is not that machine learning is used to anticipate and execute possible actions of the vehicle and the driver, but the system's legal and ethical implications. Based on the intention recognition, the system starts to prefilter the possible actions of the driver. This information is then compared with the driver's current behavior and prioritized by the system according to desirability or likelihood of a safe outcome. It should be noted that the system constantly compares its predictions with the current behavior of the driver, so it is possible for the driver to override incorrect system assumptions. This is because the system only influences the driver's steering and pedal movements and does not actively override them. Therefore, the human actions remain just that: Human.

Accordingly, there is no pre-programmed decision but rather a multi-level system of visual and auditory cues and haptic feedback leading to an improved version of the driver's behavior. By abstracting the vehicle affordances and the human intention, the system can achieve at least a safe ground state even if the driver is not available. This means that the decisions are in principle, initiated by the human driver. Therefore, this system persists as human action in a legal framework. In the case of an unresponsive driver, e.g., a medical emergency, the minimal risk state accounts in these cases.

This is, of course, only a very first draft, which remains vague regarding the exact details, and which neither deals with suitable machine learning methods nor with the implementation of the previously mentioned Bayes Theorem. An implementation in a VR experiment seems quite possible since the virtual environment allows a controlled and safe environment for testing such functionalities. Nevertheless, as mentioned in the Introduction, the real world is much more complex than a virtual environment, and the vehicle will always encounter situations for which it has not been trained. Therefore, transferring the system presented here into the real world is questionable.

Another aspect of the experiment that needs to be considered is the handling of objects and the possible replications. In the preceding experiments, we used kinematics for all dynamic objects to create a lower computational load, a stable visual input over each experiment, and facilitate changes in the scene. Due to the interactions necessary for the current experiment, we needed to access the physics engine of Unity, which does not allow the same amount of control as using kinematics for dynamic objects but provides a more realistic environment, which improves the feeling of presence. Even though there are some limitations, we still argue that we provide a functional experimental design that can test the efficiency of adaptive and cooperative HMI in critical events, where a takeover is needed.

The presented project proposal findings could help strengthen the connection between man and machine in the context of automated vehicles of level 3 and higher and thus increase the traffic safety of such systems. In addition, conclusions could be drawn from the HMI presented here about how humans and embodied agents could cooperate more generally. The presented methods of the vehicle and human affordances, together with the intention recognition as well as the likelihood calculations, could be used to foster cooperation between humans and robots in various ways.

6.3 | Concluding Remarks

Finally, it should once again be emphasized that the entire project Westdrive is thought to serve future VR research. We are more than proud to know that new projects at the IKW, namely SpaRE, EEDA, HumanA and WD Ride are based on the project Westdrive. Additionally we are happy to see that the TU Chemnitz, Korea University Seol and BASt already use our toolbox. Again I would like to express my gratitude for the wonderful years in this working group and I am excited to see the many great ideas for the continuation and extension of Project Westdrive.

7

Appendix

7.1 | LoopAR Questionnaires

The following three questionnaires are part of the LoopAR experiemental design. This accounts for the published toolkit as for the experimental setup. First Questionnaire is the System Usability scale that was used to verify the usability of the LoopAR toolkit itself. The two following questionnaires are part of the LoopAR experiment. Here we used the AVAM to later on ask for attitudes towards self-driving cars, after participants experienced one of the three experimental conditions of that experiment. All three questionnaires were created using Google Forms, handed out in Paper form.

Table 7.1: Westdrive X LoopAR Usability Questionnaire

Westdrive X LoopAR Usability Questionnaire

To validate how useful our build toolkit really is, we would like to assess our toolkit's usability with a simple, ten-item attitude Likert scale giving a global view of subjective assessments of usability. It was developed by John Brooke at Digital Equipment Corporation in the UK in 1986 as a tool to be used in usability engineering of electronic office systems.

Here, you will be asked to do specific tasks in our VR Toolbox to see if it is usable for you as our end user. The goal is to identify strengths and weaknesses in usability, so we as the developer team know what we need to improve because our goal is to provide a working VR toolkit for a broad range of users.
All tasks combined will take roughly 4 hours in total.

Additionally to the tasks, we would like to gather personal data to see who benefits the most from our toolkit. Just go through the following questionnaire. Everything else will be explained here in detail.

If you have any questions or issues, please contact mwaechter@uos.de or fnosratnezam@uos.de.

Here is the link to our GitHub repository: https://github.com/Westdrive-Workgroup/LoopAR-public

Thanks in advance for your feedback!

Team Westdrive

* Erforderlich

1. Age *

2. Gender *

Markieren Sie nur ein Oval.

- Female
- Male
- Other
- No answer

3. Education *

4. Your Field of Expertise *

5. Experience in C# *

6. Experience in Unity *

Tasks in the LoopAR toolkit

Dear tester, now it's time for you to test our toolkit. In this section, we would like to ask you to perform certain tasks. These tasks include:

- Task 1: Clone the Repository from the Git Repository to your computer https://github.com/Westdrive-Workgroup/LoopAR-public

- Task 2: Watch the tutorial videos linked to the readme file of the repo.

- Task 3: Download the right Unity version. Open the project (for example, via Unity Hub).

- Task 4: Open one scene of your choice. Change the terrain and place/delete trees, buildings, and some street properties.

- Task 5: Find a car prefab in the assets and place it onto a street.

- Task 6: Find a critical traffic event in the scene.

- Task 7: Choose one pedestrian and one car (at least) as an event object. Make paths for them to follow. Make them event objects. Get it going!

- Task 8: Change an existing critical traffic event. Use a critical traffic event prefab.

- Task 9: Find the data saving trigger and test your adjusted critical traffic event in play mode.

In the following, we will ask to perform and rate these in the following Tasks section.

7. Task 1: Issue severity

Markieren Sie nur ein Oval.

- Low: Users may experience insignificant time delays or mild frustration, but will be able to complete desired task.
- Medium: Users may experience noticeable delay or frustration, but will be able to complete the task with added effort.
- High: User will experience noticeable delay or frustration, may not be able to complete the task.

8. Task 1: Comments

9. Task 2: Issue severity

Markieren Sie nur ein Oval.

- Low: Users may experience insignificant time delays or mild frustration, but will be able to complete desired task.
- Medium: Users may experience noticeable delay or frustration, but will be able to complete the task with added effort.
- High: User will experience noticeable delay or frustration, may not be able to complete the task.

10. Task 2: Comments

11. Task 3: Issue severity

Markieren Sie nur ein Oval.

- Low: Users may experience insignificant time delays or mild frustration, but will be able to complete desired task.
- Medium: Users may experience noticeable delay or frustration, but will be able to complete the task with added effort.
- High: User will experience noticeable delay or frustration, may not be able to complete the task.

12. Task 3: Comments

13. Task 4: Issue severity

Markieren Sie nur ein Oval.

- Low: Users may experience insignificant time delays or mild frustration, but will be able to complete desired task.
- Medium: Users may experience noticeable delay or frustration, but will be able to complete the task with added effort.
- High: User will experience noticeable delay or frustration, may not be able to complete the task.

14. Task 4: Comments

15. Task 5: Issue severity

Markieren Sie nur ein Oval.

- Low: Users may experience insignificant time delays or mild frustration, but will be able to complete desired task.
- Medium: Users may experience noticeable delay or frustration, but will be able to complete the task with added effort.
- High: User will experience noticeable delay or frustration, may not be able to complete the task.

16. Task 5: Comments

17. Task 6: Issue severity

Markieren Sie nur ein Oval.

- Low: Users may experience insignificant time delays or mild frustration, but will be able to complete desired task.
- Medium: Users may experience noticeable delay or frustration, but will be able to complete the task with added effort.
- High: User will experience noticeable delay or frustration, may not be able to complete the task.

18. Task 6: Comments

19. Task 7: Issue severity

Markieren Sie nur ein Oval.

- Low: Users may experience insignificant time delays or mild frustration, but will be able to complete desired task.
- Medium: Users may experience noticeable delay or frustration, but will be able to complete the task with added effort.
- High: User will experience noticeable delay or frustration, may not be able to complete the task.

20. Task 7: Comments

21. Task 8: Issue severity

Markieren Sie nur ein Oval.

- Low: Users may experience insignificant time delays or mild frustration, but will be able to complete desired task.
- Medium: Users may experience noticeable delay or frustration, but will be able to complete the task with added effort.
- High: User will experience noticeable delay or frustration, may not be able to complete the task.

22. Task 8: Comments

23. Task 9: Issue severity

Markieren Sie nur ein Oval.

- Low: Users may experience insignificant time delays or mild frustration, but will be able to complete desired task.
- Medium: Users may experience noticeable delay or frustration, but will be able to complete the task with added effort.
- High: User will experience noticeable delay or frustration, may not be able to complete the task.

24. Task 9: Comments

25. All over comments Tasks

System Usability Scale : SUS

A simple, ten-item attitude Likert scale giving a global view of subjective assessments of usability. It was developed by John Brooke[1] at Digital Equipment Corporation in the UK in 1986 as a tool to be used in usability engineering of electronic office systems.

The usability of a system, as defined by the ISO standard ISO 9241 Part 11, can be measured only by taking into account the context of the use of the system—i.e., who is using the system, what they are using it for, and the environment in which they are using it. Furthermore, measurements of usability have several different aspects:

26. I think that I would like to use this system frequently. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Strongly disagree						Strongly agree

27. I found the system unnecessarily complex. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Strongly disagree						Strongly agree

28. I thought the system was easy to use. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Strongly disagree						Strongly agree

29. I think that I would need the support of a technical person to be able to use this system. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

30. I found the various functions in this system were well integrated. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

31. I thought there was too much inconsistency in this system. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

32. I would imagine that most people would learn to use this system very quickly. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

33. I found the system very cumbersome to use. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

34. I felt very confident using the system. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

35. I needed to learn a lot of things before I could get going with this system. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	
Strongly disagree	◯	◯	◯	◯	◯	Strongly agree

Table 7.2: LoopAR AVAM german version

AVAM LoopAR / deutsch

Die deutsche Version des Experimentalfragebogens

* Erforderlich

1. Wie alt sind Sie? *

2. Die Benutzung des Fahrzeugs würde es mir ermöglichen, mein Ziel schnell zu erreichen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu								stimme voll und ganz zu

3. Der Gebrauch des Fahrzeugs würde es mir ermöglichen, mein Ziel kostengünstig zu erreichen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu								stimme voll und ganz zu

4. Das Fahrzeug würde es mir ermöglichen, mein Ziel sicher zu erreichen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu								stimme voll und ganz zu

5. Ich fände das Fahrzeug leicht zu bedienen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu	◯	◯	◯	◯	◯	◯	◯	stimme voll und ganz zu

6. Meine Interaktion mit dem Fahrzeug wäre einfach und verständlich. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu	◯	◯	◯	◯	◯	◯	◯	stimme voll und ganz zu

7. Es wäre für mich ein Leichtes, den Umgang mit dem Fahrzeug zu erlernen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu	◯	◯	◯	◯	◯	◯	◯	stimme voll und ganz zu

8. Ich wäre stolz darauf, das Fahrzeug den Menschen zu zeigen, die mir nahe stehen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu	◯	◯	◯	◯	◯	◯	◯	stimme voll und ganz zu

9. Ich würde mich eher geneigt fühlen, das Fahrzeug zu benutzen, wenn es von anderen Menschen häufig benutzt würde. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu	◯	◯	◯	◯	◯	◯	◯	stimme voll und ganz zu

10. Ich würde es vorziehen, das Fahrzeug auch mit anderen Passagieren im Fahrzeug zu benutzen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu	◯	◯	◯	◯	◯	◯	◯	stimme voll und ganz z

11. Ich hätte eine angemessene Kontrolle über die Reise zu meinem Ziel. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu	◯	◯	◯	◯	◯	◯	◯	stimme voll und ganz z

12. Ich verfüge über die erforderlichen Kenntnisse zur Benutzung des Fahrzeugs. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu	◯	◯	◯	◯	◯	◯	◯	stimme voll und ganz z

13. Das Fahrzeug und die Infrastruktur, die zur Nutzung des Fahrzeugs erforderlich sind, sind praktisch vorhanden. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu	◯	◯	◯	◯	◯	◯	◯	stimme voll und ganz z

14. Es wäre eine gute Idee, das Fahrzeug zu benutzen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu	◯	◯	◯	◯	◯	◯	◯	stimme voll und ganz :

15. Das Fahrzeug würde das Fahren interessanter machen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu	◯	◯	◯	◯	◯	◯	◯	stimme voll und ganz :

16. Es würde Spaß machen, das Fahrzeug zu benutzen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu	◯	◯	◯	◯	◯	◯	◯	stimme voll und ganz :

17. Ich könnte mein Ziel mit dem Fahrzeug erreichen, wenn ich nur die integrierte Anleitung zur Hilfestellung hätte. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu	◯	◯	◯	◯	◯	◯	◯	stimme voll und ganz :

18. Ich könnte mein Ziel mit dem Fahrzeug erreichen, wenn ich keine Hilfe hätte. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu	◯	◯	◯	◯	◯	◯	◯	stimme voll und ganz :

19. Ich könnte mein Ziel mit dem Fahrzeug erreichen, wenn es jemanden gäbe, der mir helfen könnte. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu								stimme voll und ganz z

20. Ich hätte Bedenken, das Fahrzeug zu benutzen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu								stimme voll und ganz z

21. Das Fahrzeug wäre für mich irgendwie beängstigend. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu								stimme voll und ganz z

22. Ich fürchte, dass ich das Fahrzeug nicht verstehen würde. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu								stimme voll und ganz z

23. Angenommen, ich hätte Zugang zu dem Fahrzeug, dann würde ich es auch benutzen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu								stimme voll und ganz z

24. Wenn mir das Fahrzeug zur Verfügung steht, plane ich, es zu beschaffen und zu nutzen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu								stimme voll und ganz z

25. Ich glaube, dass es gefährlich wäre, das Fahrzeug zu benutzen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu								stimme voll und ganz z

26. Ich würde mich bei der Benutzung des Fahrzeugs sicher fühlen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu								stimme voll und ganz z

27. Ich würde dem Fahrzeug vertrauen. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
stimme überhaupt nicht zu								stimme voll und ganz z

28. Wie wichtig wären Ihre Augen bei der Benutzung des Fahrzeugs? *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
überhaupt nicht wichtig								sehr wichtig

29. Wie wichtig wären Ihre Hände bei der Benutzung des Fahrzeugs? *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
überhaupt nicht wichtig	⬭	⬭	⬭	⬭	⬭	⬭	⬭	sehr wichtig

30. Wie wichtig wären Ihre Füße bei der Benutzung des Fahrzeugs? *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
überhaupt nicht wichtig	⬭	⬭	⬭	⬭	⬭	⬭	⬭	sehr wichtig

31. Mit welchem Geschlecht identifizieren Sie sich *

Markieren Sie nur ein Oval.

- Weiblich
- Männlich
- Intersex
- Sonstiges: ____________________

32. Wie viele Stunden fahren Sie pro Woche Auto? *

____

33. Wofür nutzen Sie das Auto vorrangig? *

____

34. Wie viele Stunden pro Woche spielen Sie Computerspiele? *

____

35. Wie oft nutzen Sie virtuelle Realitäten? *

Markieren Sie nur ein Oval.

- Noch Nie
- Selten
- Manchmal
- Oft

36. Besitzen Sie einen Führerschein?

Markieren Sie nur ein Oval.

- Ja
- Nein

Table 7.3: LoopAR AVAM english version

AVAM LoopAR / english

This gonna be the english version of the AVAM, that we will use at the LoopAR Experiment

* Erforderlich

1. What is your age? *

2. What gender do you identify with? *

Markieren Sie nur ein Oval.

- Female
- Male
- Intersex
- Sonstiges: ______

3. Using the vehicle would enable me to reach my destination quickly. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

4. Using the vehicle would enable me to reach my destination cost efficiently. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

5. Using the vehicle would enable me to reach my destination safely. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

6. I would find the vehicle easy to use. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

7. My interaction with the vehicle would be clear and understandable. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

8. It would be easy for me to learn to use the vehicle. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

9. I would be proud to show the vehicle to people who are close to me. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	strongly agree

10. I would feel more inclined to use the vehicle if it was widely used by others. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	strongly agree

11. I would prefer to use the vehicle with other passengers in the vehicle as well. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	strongly agree

12. I would have adequate control over the journey to my destination. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	strongly agree

13. I have the knowledge necessary to use the vehicle. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

14. The vehicle and infrastructure necessary to use the vehicle are practically feasible. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

15. Using the vehicle would be a good idea. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

16. The vehicle would make driving more interesting. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

17. Using the vehicle would be fun. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	strongly agree

18. I could reach my destination using the vehicle if I had just the built-in instruction for assistance. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	strongly agree

19. I could reach my destination using the vehicle if I had no assistance. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	strongly agree

20. I could reach my destination using the vehicle if there was someone who could help me. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree	⬭	⬭	⬭	⬭	⬭	⬭	⬭	strongly agree

21. I would have concerns about using the vehicle. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

22. The vehicle would be somewhat frightening to me. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

23. I am afraid that I would not understand the vehicle. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

24. Given that I had access to the vehicle, I predict that I would use it. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

25. If the vehicle becomes available to me, I plan to obtain and use it. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

26. I believe that using the vehicle would be dangerous. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

27. I would feel safe while using the vehicle. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

28. I would trust the vehicle. *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
strongly disagree								strongly agree

29. How important would your eyes be when using the vehicle? *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
Not important at all	◯	◯	◯	◯	◯	◯	◯	very important

30. How important would your hands be when using the vehicle? *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
Not important at all	◯	◯	◯	◯	◯	◯	◯	very important

31. How important would your feet be when using the vehicle? *

Markieren Sie nur ein Oval.

	1	2	3	4	5	6	7	
Not important at all	◯	◯	◯	◯	◯	◯	◯	very important

32. How many hours do you drive per week? *

____

33. For what reasons do you typically drive? *

____

34. How many hours per weeks do you play video games? *

____

35. How often do you use virtual reality? *

Markieren Sie nur ein Oval.

- Never
- Occasionally
- Sometimes
- Often

7.2 | Motor City: Supplementary Material

The Supplementary Material for this article can be found at:
https://www.frontiersin.org/articles/10.3389/fict.2020.00001/full#supplementary-material

Westdrive recommended system specs:
GPU: NVidia GeForce GTX 1080 ti, equivalent or better.
CPU: Intel(R) Xenon ® E5-1607 v4, equivalent or better.
RAM: 32 GB
Video Output: HDMI 1.4, DisplayPort 1.2 or newer.
Operating System: Windows 8.1 or later, Windows 10.

Character Creation: To create new characters for the city it needs two external tools. One is Fuse CC from Adobe and the other is the Mixamo website. In Fuse CC, a free 3D design program by adobe, it is possible to create figures according to your own imagination. This created model can then be uploaded to the program at Mixamo, which can automatically create animations (Aguiar et al., 2014). From this website, finished models with animations can be downloaded, which then only need to be implemented into the project.

Unity® 3D learning: www.unity.com/learn.

Online Character animation: www.mixamo.com.

7.3 | Talking Cars-doubtful users: Supplementary Materials

Methods

To perform the Manova and it's follow-up data has been prepared for transfer from python pandas CSV format by replacing categorical labels to numerical values. Afterward the following statement has been used in SPSS in order to calculate MANOVA and follow-up LDA as well as effect sizes.

MANOVA Intention Usefulness Ease Trust by AgeGroup(0,4) Gender(0,1) Condition(0,2)
/DISCRIM=STAN RAW CORR
/PRINT=SIGNIF(MULTIV,UNIV,EIGEN,DIMENR)
/DESIGN AgeGroup, Gender, Condition, AgeGroup by Gender

COMPUTE Super_Condition = $(-.03420 * Intention) + (.03247 * Usefulness) + (-.01176 * Ease) + (.01498 * Trust)$ EXECUTE.

COMPUTE Super_Gender = $(-.01026 * Intention) + (-.00212 * Usefulness) + (-.02059 * Ease) + (-.00629 * Trust)$. EXECUTE.

COMPUTE Super_AgeGroup = $(-.00850 * Intention) + (-.00647 * Usefulness) + (-.01570 * Ease) + (-.01029 * Trust)$. EXECUTE.

COMPUTE Super_AgeGroup_2 = $(.01947 * Intention) + (.01436 * Usefulness) + (-.00371 * Ease) + (-.03451 * Trust)$. EXECUTE.

COMPUTE Super_AgeGroup_Gender = $(-.02251 * Intention) + (.00903 * Usefulness) + (-.02492 * Ease) + (-.01862 * Trust)$. EXECUTE.

Category	Question	slider	Possible Answer
Intention to use 1	Ich würde so ein Auto im echten Verkehr nutzen	y	0-100
Perceived Usefulness 4	Ich denke, dass das gesehene Fahrzeug in meinem Alltag nützlich wäre.	y	0-100
Perceived Ease of Use 2	Die Nutzung von so einem Fahrzeug würde mir leicht fallen.	y	0-100
Trust	Ich habe mich während der Fahrt sicher gefühlt .	y	0-100
Sex	Bitte benenne dein Geschlecht.	n	männlich weiblich intersex keine Angabe
Age	Dein Alter:	n	numpad
Aviophobia	Hast du Flugangst?	n	Ja/Nein
Driving Frequency	Wie viele Jahre fährst du schon regelmäßig?	n	numpad
Playing Hours	Wie viele Stunden in der Woche spielst du ungefähr Videospiele?	n	numpad
VR Playing Frequency	Wie oft hast du schon Virtual Reality genutzt?	n	Noch nie Ein mal unter 10 mehr als 10 mal

Table 7.4: Simplified TAM questionnaire for the MS Wissenschaft experimental setup. Here the initial TAM 2 was reduced to one question per Item

Tables for all calculated effect

age group	below 20	20 - 40	40 - 60	60 - 80	above 80
below 20	0	0.40	0.50	0.43	0.17 / 0.2
20 - 40	0.40	0	0.10	0.01	0.18 / 0.20
40 - 60	0.50	0.10	0	0.08	0.28
60 - 80	0.43	0.01	0.08	0	0.20 / 0.22
above 80	0.17 / 0.2	0.18 / 0.20	0.28	0.20 / 0.22	0

Table 7.5: Effect Sizes between different age groups on Intention to use, perceived usefulness, perceived ease of use and trust. Numbers in the table present the Cohen's D and in the case of difference Hedges G

condition	AVAS	RadioTalk	TaxiDriver
AVAS	0	0.05	0.11
RadioTalk	0.99	0	0.06
TaxiDriver	0.11	0.06	0

Table 7.6: Effect Sizes between different condition on Intention to use, perceived usefulness, perceived ease of use and trust. Numbers in the table present the Cohen's D and in the case of difference Hedges G

Gender/Age Group	Male / below 20	Male / 20 - 40	Male / 40 - 60	Male / 60 - 80	Male / 80+
Male / below 20	0	0.30	0.36	0.39	0.28 - 0.32
Male / 20 - 40	0.30	0	0.06	0.07	0
Male / 40 - 60	0.36	0.06	0	0.01	0.05
Male / 60 - 80	0.39	0.07	0.01	0	0.06
Male / 80+	0.28 - 0.32	0	0.05	0.06	0
Female / below 20	0.31	0.01	0.05	0.06	0.04
Female / 20 - 40	0.73	0.42	0.36	0.35	0.39 - 0.41
Female / 40 - 60	0.90	0.58	0.51	0.51	0.53 - 0.56
Female / 60 - 80	0.78 - 0.80	0.46	0.40	0.40	0.42 - 0.44
Female / 80+	0.29 - 0-33	0	0.03	0.05	0.01

Table 7.7: Effect Sizes between different combination of gender and age group on Intention to use, perceived usefulness, perceived ease of use and trust. Numbers in the table present the Cohen's D and in the case of difference Hedges G

Gender/Age Group	Female / below 20	Female / 20 - 40	Female / 40 - 60	Female / 60 - 80	Female / 80+
Male / below 20	0.31	0.73	0.9	0.78 - 0.80	0.29 - 0-33
Male / 20 - 40	0.01	0.42	0.58	0.46	0
Male / 40 - 60	0.05	0.36	0.51	0.40	0.03
Male / 60 - 80	0.06	0.35	0.51	0.40	0.05
Male / 80+	0.04	0.39 - 0.41	0.53 - 0.56	0.42 - 0.44	0.01
Female / below 20	0	0.42	0.57	0.46	0.01
Female / 20 - 40	0.42	0	0.15	0.03	0.37 - 0.40
Female / 40 - 60	0.57	0.15	0	0.11	0.51 - 0.55
Female / 60 - 80	0.46	0.03	0.11	0	0.41 - 0.44
Female / 80+	0.01	0.37 - 0.40	0.51 - 0.55	0.41 - 0.44	0

Table 7.8: Effect Sizes between different combination of gender and age group on Intention to use, perceived usefulness, perceived ease of use and trust. Numbers in the table present the Cohen's D and in the case of difference Hedges G

Disclaimer

All experiments reported in this thesis are conform with the Declaration of Helsinki. Additionally, all experiments have been approved by the ethics committee of the University of Osnabrück. All measurements within the faculty in 2020 and 2021 were carried out with a hygiene concept approved by the university.

I hereby confirm that I wrote this thesis independently and that I have not made use of resources other than indicated in this thesis. I guarantee that I significantly contributed to all materials used. This thesis was not published, expect the parts indicated above. Any of these publications, not this thesis itself have been used to fulfill any other examination requirements.

Publication List

Journal Articles

Faulhaber, A. K.†, Dittmer, A.†, Blind, F.†, Wächter, M. A.†', Timm, S.†, Sütfeld, L. R., Pipa, G. & König, P. (2019). *Human decisions in moral dilemmas are largely described by utilitarianism: Virtual car driving study provides guidelines for autonomous driving vehicles*. Science and engineering ethics, 25(2), 399-418.

Nezami, F. N.†, Wächter, M. A.†', Pipa, G., & König, P. (2020). *Project Westdrive: Unity city with self-driving cars and pedestrians for virtual reality studies*. Frontiers in ICT, 7, 1.

Nezami, F. N.†, Wächter, M. A.†', Maleki, N., Spaniol, P., Kühne, L. M., Haas, A., Pingel, J. M., Tiemann, L., Nienhaus, F., Keller, L., König, S. U, König, P. & Pipa, G. (2021). *Westdrive X LoopAR: An Open-Access Virtual Reality Project in Unity for Evaluating User Interaction Methods during Takeover Requests*. Sensors, 21(5), 1879.

Derakhshan S.†, Nezami, F.N.†', Wächter, M.A.†, Keshava, A.,Vidal De Palol, M., Lukanov, H., Czeszumski, A., Pipa, G., König, P. *Talking cars, doubtful users - a population study in virtual reality* IEEE Transactions on Human-Machine Systems.

† shared first authorship
' corresponding author

Preprint Articles

Nezami, F. N.†, Wächter, M. A.†, Maleki, N., Spaniol, P., Kühne, L. M., Haas, A., Pingel, J. M., Tiemann, L., Nienhaus, F., Keller, L., König, S. U, König, P. & Pipa, G. (2020, May 25). *From Interaction to Cooperation: a new approach for human-machine interaction research for closing the out-of-the-loop unfamiliarity.* psyarxiv 10.31234/osf.io/7jg3c

Nezami, F. N.†, Wächter, M. A.†, Keshava, A. Lukanov, H. Vidal de Palol, M., Pipa, G. & König, P. *Talking Cars, doubtful users - a population study in virtual reality* psycrxiv 10.31234/osf.io/bsjy6

Invited Talks

Wächter, M. A., Nezami, F.N., (December, 2018) *Innovation and Regulation regarding Self-Driving Cars*, Workshop for the profile line 1 "Digitale Gesellschaft - Innovation - Regulierung", Osnabrück.

Wächter, M. A. (January, 2019) *Akzeptanz und dem Vertrauen in autonome Fahrzeuge im Rahmen von skalierbaren Experimenten in der virtuellen Realität*, Ringvorlesung des GK Vertrauen und Akzepanz in erweiterten virtuellen Arbeitswelten, Osnabrück

Wächter, M.A., Nezami, F.N., (April, 2019) *Virtuelle Realitäten in der Verkehrsforschung: Project Westdrive der Uni Osnabrück*, Forschung Made in Niedersachen, Braunschweig

Wächter, M. A. (May, 2019) *Steig ein - heute Fährt dein Auto selbst! Simulatorstudien zur Akzeptanz in der virtuellen Realität*, Eröffnung der MS Wissenschaft 2019, Berlin

Wächter, M. A., Nezami, F. N., (June, 2019) *Projekt Westdrive: Mensch-Maschine Interaktion in der virtuellen Realität*, BMBF KarliczekImpulse Wissenschaftsjahr 2019, Berlin

Wächter, M. A., Nezami, F. N. (July, 2019) *Selbsterklärende künstliche Intelligenz und virtuelle Autos: Ein Forschungprojekt der Uni Osnabrück und der Stiftung Stahlwerk Georgsmarienhütte*, IdeenExpo 2019, Hannover

Wächter, M. A. (October, 2019) *Projekt Westdrive: groß angelegte VR Studien in der Mensch-Maschine Interaktion* 13th Doctoral Workshop of the Section of Traffic Psychology, Münster

Wächter, M. A., Pipa, G. (February, 2020) *Künstliche Intelligenz in Niedersachsen – Chancen und Herausforderungen*, KI Ministerpräsidentenreise, Osnabrück

Wächter, M. A., Pipa, G. (February, 2020) *Potenziale der Künstlichen Intelligenz*, Wirtschaftsvereinigung Grafschaft Bentheim, Bad Bentheim

Wächter, M. A., Nezami, F. N. (July, 2021) *Virtual Reality - Von menschlichen Verhalten bis zur Städteplanung*, OSNAhack 2021, Osnabrück

Wächter, M. A. (September, 2021) *Vertrauen und Akzeptanz in KI Agenten: Beispiele aus der virtuellen Realität zum Thema selbstfahrende Fahrzeuge*, MeetUp – Wissenschaft trifft Wirtschaft, Osnabrück

Bibliography

Abdi, H., & Williams, L. J. (2010). Tukey's honestly significant difference (HSD) test. *Encyclopedia of research design*, *3*(1), 1–5.

Abe, G., Itoh, M., & Yamamura, T. (2011). Effective and acceptable forward collision warning systems based on relationships between car-following behaviour and reaction to deceleration of lead vehicle. *Human modelling in assisted transportation* (pp. 155–164). Springer.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE access*, *6*, 52138–52160.

Addagarla, S. K., & Amalanathan, A. (2020). Probabilistic unsupervised machine learning approach for a similar image recommender system for e-commerce. *Symmetry*, *12*(11), 1783. Retrieved August 19, 2021, from https://www.mdpi.com/2073-8994/12/11/1783

Aguiar, E. D., Gambaretto, E., & Corazza, S. (2014, June 12). *Web platform for interactive design, synthesis and delivery of 3d character motion data* (U.S. pat. 20140160116A1). Retrieved June 3, 2021, from https://patents.google.com/patent/US20140160116A1/en

Ahmad, S. (2020, February 14). *HTC Vive: Vor- und Nachteile des VR-Headsets World of VR* [World of VR]. Retrieved June 3, 2021, from https://worldofvr.net/htc-vive/

Allahyari, T., Saraji, G. N., Adi, J., Hosseini, M., Iravani, M., Younesian, M., & Kass, S. J. (2008). Cognitive failures, driving errors and driving accidents. *International journal of occupational safety and ergonomics*, *14*(2), 149–158.

Alt, R., Leimeister, J. M., Priemuth, T., Sachse, S., Urbach, N., & Wunderlich, N. (2020). Software-defined business. *Business & Information Systems Engineering*, *62*(6), 609–621.

Altendorf, E., Schütz, R., Canpolat, Y., Weßel, G., & Flemisch, F. (2017). A study on the human and the automation in automated driving: Getting to know each other. *International Conference on Applied Human Factors and Ergonomics*, 672–683.

Anthes, C., García-Hernández, R. J., Wiedemann, M., & Kranzlmüller, D. (2016). State of the art of virtual reality technology. *2016 IEEE Aerospace Conference*, 1–19.

Anwar, M. D. (2019, May). *How virtual reality works*. Retrieved November 18, 2021, from http://www.laboreldi.or.id/2019/05/bagaimana-virtual-reality-bekerja.html

ArcGIS. (2013). 3d modeling software for urban environments. *ESRI*. Retrieved June 3, 2021, from https://desktop.arcgis.com/de/cityengine/

Asimov, I. (2013, June 6). *I, robot* (1. Edition). HarperVoyager.

Ateş, C., Kaymaz, Ö., Kale, H. E., & Tekindal, M. A. (2019). Comparison of test statistics of nonnormal and unbalanced samples for multivariate analysis of variance in terms of type-i error rates. *Comput. Math. Methods Med.*, *2019*, 2173638.

Audi AG. (2021). *Audi technology portal - audi a8 – audi AI traffic jam pilot* [Audi technology portal]. Retrieved October 21, 2021, from https://www.audi-technology-portal.de/en/electrics-electronics/driver-assistant-systems/audi-a8-audi-ai-traffic-jam-pilot

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

Baltes, M. M., Wahl, H. W., & Schmid-Furstoss, U. (1990). The daily life of elderly germans: Activity patterns, personal control, and functional health. *J. Gerontol.*, *45*(4), P173–9.

Bashiri, A., Ghazisaeedi, M., & Shahmoradi, L. (2017). The opportunities of virtual reality in the rehabilitation of children with attention deficit hyperactivity disorder: A literature review. *Korean Journal of Pediatrics*, *60*(11), 337.

Bechtel, M. G., Mcellhiney, E., Kim, M., & Yun, H. (2018). DeepPicar: A low-cost deep neural network-based autonomous car. *2018 IEEE 24th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)*, 11–21.

Belanche, D., Casaló, L. V., & Flavián, C. (2012). Integrating trust and personal values into the technology acceptance model: The case of e-government services adoption. *Cuadernos de Economía y Dirección de la Empresa*, *15*(4), 192–204.

Bengler, K., Rettenmaier, M., Fritz, N., & Feierle, A. (2020). From HMI to HMIs: Towards an HMI framework for automated driving. *Information*, *11*(2), 61.

Benleulmi, A. Z., & Blecker, T. (2017). Investigating the factors influencing the acceptance of fully autonomous cars. *Digitalization in Supply Chain Management and Logistics: Smart and Digital Solutions for an Industry 4.0 Environment. Proceedings of the Hamburg International Conference of Logistics (HICL), Vol. 23*, 99–115.

Bergmann, L. T., Schlicht, L., Meixner, C., König, P., Pipa, G., Boshammer, S., & Stephan, A. (2018). Autonomous vehicles require socio-political acceptance—an empirical and philosophical perspective on the problem of moral decision making. *Frontiers in Behavioral Neuroscience*, *12*, 31.

Bischof, W. F., & Boulanger, P. (2003). Spatial navigation in virtual reality environments: An EEG analysis. *CyberPsychology & Behavior*, *6*(5), 487–495.

Bjørner, T. (2019). Driving pleasure and perceptions of the transition from no automation to full self-driving automation. *Appl. mobilities*, *4*(3), 257–272.

Blackman, G. (2018). Real commitment for lidar from automotive OEMs. *Imaging and Machine Vision Europe*, (87), 10–12.

Blender Foundation. (1995). *Blender.org - home of the blender project - free and open 3d creation software* [Blender.org]. Retrieved June 10, 2021, from https://www.blender.org/

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2015). Autonomous vehicles need experimental ethics: Are we ready for utilitarian cars? *ArXiv*. Retrieved June 1, 2021, from https://arxiv.org/pdf/1510.03346v1.pdf

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573–1576.

Bose, I., & Mahapatra, R. K. (2001). Business data mining — a machine learning perspective. *Information & Management*, *39*(3), 211–225.

Botica, N., Martins, M., Ribeiro, M. d. C. F., & Magalhães, F. (2015). *3d representation of the urban evolution of braga using the cityengine tool*. Vest-Agder-museet. Retrieved June 3, 2021, from http://repositorium.sdum.uminho.pt/

Bovet, A., & Makse, H. A. (2019). Influence of fake news in twitter during the 2016 US presidential election. *Nature Communications*, *10*(1), 7.

Bundesministerium für Digitales und Verkehr. (2017, June 20). *Bericht der ethikkommission: Automatisiertes und vernetztes fahren* [Bundesministerium für verkehr und digitale infrastruktur]. Retrieved November 21, 2021, from https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.html

Bundesministeriums der Justiz und für Verbraucherschutz. (2013). *Straßenverkehrsordnung vom 6. märz 2013 (BGBl. i s. 367), die zuletzt durch artikel 13 des gesetzes vom 12.juli 2021 (BGBl. i s. 3091) geändert worden ist.* Retrieved October 14, 2021, from https://www.gesetze-im-internet.de/stvo_2013/StVO.pdf

Burke, R. R. (2018). Virtual reality for marketing research. *Innovative research methodologies in management* (pp. 63–82). Springer.

Butt, A. L., Kardong-Edgren, S., & Ellertson, A. (2018). Using game-based virtual reality with haptics for skill acquisition. *Clinical Simulation in Nursing*, *16*, 25–32.

Campbell, M., Egerstedt, M., How, J. P., & Murray, R. M. (2010). Autonomous driving in urban environments: Approaches, lessons and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1928), 4649–4672.

Campisi, T., Severino, A., Al-Rashid, M. A., & Pau, G. (2021). The development of the smart cities in the connected and autonomous vehicles (CAVs) era: From mobility patterns to scaling in cities. *Infrastructures*, *6*(7), 100.

Carlini, N., & Wagner, D. (2017). Towards evaluating the robustness of neural networks. *2017 IEEE Symposium on Security and Privacy (SP)*, 39–57.

Carranza-García, M., Torres-Mateo, J., Lara-Benítez, P., & García-Gutiérrez, J. (2021). On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data. *Remote Sensing*, *13*(1), 89.

Castelli, L., Latini Corazzini, L., & Geminiani, G. C. (2008). Spatial navigation in large-scale virtual environments: Gender differences in survey tasks. *Computers in Human Behavior*, *24*(4), 1643–1667.

Check Hayden, E. (2014). The automated lab. *Nature*, *516*(7529), 131–132.

Chehri, A., & Mouftah, H. T. (2019). Autonomous vehicles in the sustainable cities, the beginning of a green adventure. *Sustainable Cities and Society*, *51*, 101751.

Chen, F., Shang, Y., Hu, J., & Xu, B. (2020). Few features attack to fool machine learning models through mask-based GAN. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–7.

Chen, J.-F., Hsieh, H.-N., & Do, Q. (2014). Forecasting hoabinh reservoir's incoming flow: An application of neural networks with the cuckoo search algorithm. *Information*, *5*(4), 570–586.

Chen, K., & Chan, A. H. S. (2011). A review of technology acceptance by older adults. *Gerontechnology*, *10*(1), 1–12.

Chialastri, A. (2012). Automation in aviation. *Automation*. IntechOpen.

Choi, B. C., & Pak, A. W. (2005). Peer reviewed: A catalog of biases in questionnaires. *Preventing chronic disease*, *2*(1).

Choi, J.-w., Curry, R. E., & Elkaim, G. H. (2010). Continuous curvature path generation based on bézier curves for autonomous vehicles. *International Journal of Applied Mathematics*, *40*(2).

Clark, A. (2006). Vision as dance? three challenges for sensorimotor contingency theory.

Clay, V., König, P., & König, S. U. (2019). Eye tracking in virtual reality. *Journal of Eye Movement Research*, *12*(1).

Coelingh, E., & Ekmark, J. (2019). Selbstfahrende fahrzeuge revolutionieren das verkehrswesen. *ATZelektronik*, *14*(9), 16–21.

Cohen, I. G., Evgeniou, T., Gerke, S., & Minssen, T. (2020). The european artificial intelligence strategy: Implications and challenges for digital health. *The Lancet Digital Health*, *2*(7), e376–e379.

Connors, J., Graham, S., & Mailloux, L. (2018). Cyber synthetic modeling for vehicle-to-vehicle applications. *International Conference on Cyber Warfare and Security*, 594–XI.

Cruden. (2018). *Automotive driving simulators* [Cruden.com]. Retrieved June 10, 2021, from https://www.cruden.com/automotive-driving-simulators/

Cusumano, M. A. (2020). Self-driving vehicle technology: Progress and promises. *Communications of the ACM*, *63*(10), 20–22.

Cybenko, A. K., & Cybenko, G. (2018). AI and fake news. *IEEE Intelligent Systems*, *33*(5), 1–5.

Cysneiros, L. M., Raffi, M., & Sampaio do Prado Leite, J. C. (2018). Software transparency as a key requirement for self-driving cars. *2018 IEEE 26th International Requirements Engineering Conference (RE)*, 382–387.

Czaja, S. J., Sharit, J., Charness, N., Fisk, A. D., & Rogers, W. (2001). The center for research and education on aging and technology enhancement (CREATE): A program to enhance technology for older adults. *Gerontechnology*, *1*(1), 50–59.

Dajsuren, Y., & van den Brand, M. (2019). Automotive software engineering: Past, present, and future. In Y. Dajsuren & M. van den Brand (Eds.), *Automotive systems and software engineering: State of the art and future trends* (pp. 3–8). Springer International Publishing.

Dale, R. (2021). GPT-3: What's it good for? *Natural Language Engineering*, *27*(1), 113–118.

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94–98.

Davis, F. D., & Venkatesh, V. (1996). A critical assessment of potential measurement biases in the technology acceptance model: Three experiments [Elsevier BV]. *Int. J. Hum. Comput. Stud.*, *45*(1), 19–45.

de la Rosa, S., & Breidt, M. (2018). Virtual reality: A new track in psychological research. *British Journal of Psychology*, *109*(3), 427–430.

Demel, H., & Hemming, H. (1989). ABS and ASR for passenger cars -coals and limits, 890834.

Dietsche, K.-H., & Kuhlgatz, D. (2014). History of the automobile. In K. Reif (Ed.), *Fundamentals of automotive and engine technology* (pp. 1–7). Springer Fachmedien Wiesbaden.

Dix, A., Helmert, J. R., Wagner, T., & Pannasch, S. (2021). Autonom und unfallfrei: Betrachtungen zur rolle der technischen aufsicht im kontext des autonomen fahrens [PsyArXiv Preprint]. Retrieved September 7, 2021, from https://psyarxiv.com/g2upb/

Dogan, E., Honnêt, V., Masfrand, S., & Guillaume, A. (2019). Effects of non-driving-related tasks on takeover performance in different takeover situations in conditionally automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *62*, 494–504.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *preprint arXiv:1702.08608*. Retrieved October 6, 2021, from https://arxiv.org/abs/1702.08608

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017). CARLA: An open urban driving simulator. *Conference on Robot Learning*, 1–16. Retrieved June 10, 2021, from http://proceedings.mlr.press/v78/dosovitskiy17a.html

Du, N., Haspiel, J., Zhang, Q., Tilbury, D., Pradhan, A. K., Yang, X. J., & Robert, L. P., Jr. (2019). Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload [Elsevier BV]. *Transp. Res. Part C: Emerg. Technol.*, *104*, 428–442.

Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer.

Dudley, S. E., & Wegrich, K. (2016). The role of transparency in regulatory governance: Comparing US and EU regulatory systems. *Journal of Risk Research*, *19*(9), 1141–1157.

Ebers, M. (2021). Regulating explainable AI in the european union. an overview of the current legal framework(s). *SSRN Electronic Journal*.

Eby, D. W., Molnar, L. J., Zhang, L., St. Louis, R. M., Zanier, N., Kostyniuk, L. P., & Stanciu, S. (2016). Use, perceptions, and benefits of automotive technologies among aging drivers. *Injury Epidemiology*, *3*(1), 28.

Ehinger, B. V., Groß, K., Ibs, I., & König, P. (2019). A new comprehensive eye-tracking test battery concurrently evaluating the pupil labs glasses and the EyeLink 1000 [PeerJ Inc.]. *PeerJ*, *7*, e7086.

Ekstrom, A. D., & Isham, E. A. (2017). Human spatial navigation: Representations across dimensions and scales. *Current Opinion in Behavioral Sciences*, *17*, 84–89.

Endsley, M. R., Bolte, B., & Jones, D. G. (2003, July 17). *Designing for situation awareness* (0th ed.). CRC Press.

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

Engel, A. K., Maye, A., Kurthen, M., & König, P. (2013). Where's the action? the pragmatic turn in cognitive science. *Trends in cognitive sciences*, *17*(5), 202–209.

Engström, J., & Hollnagel, E. (2007). A general conceptual framework for modelling behavioural effects of driver support functions. In P. C. Cacciabue (Ed.), *Modelling driver behaviour in automotive environments* (pp. 61–84). Springer London.

Epic Games Inc. (2021). *Unreal engine | the most powerful real-time 3d creation tool* [Unreal 5 game engine]. Retrieved October 15, 2021, from https://www.unrealengine.com/en-US/

Epstein, R. A., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: Spatial navigation and beyond. *Nature Neuroscience*, *20*(11), 1504–1513.

EU commission. (2020). *WHITE PAPER on artificial intelligence - a european approach to excellence and trust.* Retrieved August 21, 2021, from https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Facebook Technologies LLC. (2021). *Oculus quest 2: Our most advanced new all-in-one VR headset | oculus* [Oculus.com]. Retrieved October 18, 2021, from https://www.oculus.com/quest-2/

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations [Elsevier]. *Transportation Research Part A: Policy and Practice*, *77*, 167–181.

Faisal, A. (2017). Computer science: Visionary of virtual reality. *Nature*, *551*(7680), 298–299.

Fang, Y., Nakashima, R., Matsumiya, K., Kuriki, I., & Shioiri, S. (2015). Eye-head coordination for visual cognitive processing (Z. Kapoula, Ed.). *PLOS ONE*, *10*(3), e0121035.

Faulhaber, A. K., Dittmer, A., Blind, F., Wächter, M. A., Timm, S., Sütfeld, L. R., Stephan, A., Pipa, G., & König, P. (2019). Human decisions in moral dilemmas are largely described by utilitarianism: Virtual car driving study provides guidelines for autonomous driving vehicles. *Science and Engineering Ethics*, *25*(2), 399–418.

Finch, H., & French, B. (2013). A monte carlo comparison of robust MANOVA test statistics [Wayne State University Library System]. *J. Mod. Appl. Stat. Methods*, *12*(2), 35–81.

Foot, P. (1967). *The problems of abortion and the doctrine of the double effect* [Georgetown University Library]. Retrieved June 1, 2021, from https://repository.library.georgetown.edu/handle/10822/762732

Forrester, J. W. (1971). Counterintuitive behavior of social systems. *Technological Forecasting and Social Change*, *3*, 1–22.

Forster, Y., Naujoks, F., & Neukum, A. (2017). Increasing anthropomorphism and trust in automated driving functions by adding speech output. *2017 IEEE Intelligent Vehicles Symposium (IV)*, 365–372.

Freeman, D., Haselton, P., Freeman, J., Spanlang, B., Kishore, S., Albery, E., Denne, M., Brown, P., Slater, M., & Nickless, A. (2018). Automated psychological therapy using immersive virtual reality for treatment of fear of heights: A single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry*, *5*(8), 625–632.

Gaupp, F., & Katzenbach, P. (2019, July). *Automated driving through systems engineering* [Unity Technologies]. Retrieved October 28, 2021, from https://www.unity.de/en/services/systems-engineering/automated-driving-through-systems-engineering/

Georgevici, A. I., & Terblanche, M. (2019). Neural networks and deep learning: A brief introduction. *Intensive Care Medicine*, *45*(5), 712–714.

Gerdes, J. C., & Thornton, S. M. (2015). Implementable Ethics for Autonomous Vehicles. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.), *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte* (pp. 87–102). Springer.

Gessner, D. (2020, June 22). *Experts say we're decades from fully autonomous cars. here's why.* [Businessinsider.com]. Retrieved August 26, 2021, from https://www.businessinsider.com/self-driving-cars-fully-autonomous-vehicles-future-prediction-timeline-2019-8

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, *521*(7553), 452–459.

Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B., & Spaulding, R. (2021). Attachment and trust in artificial intelligence. *Computers in Human Behavior*, *115*, 106607.

Gillmore, S. C., & Tenhundfeld, N. L. (2020). The good, the bad, and the ugly: Evaluating tesla's human factors in the wild west of self-driving cars. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 67–71.

Gogoll, J., & Müller, J. F. (2017). Autonomous cars: In favor of a mandatory ethics setting. *Science and Engineering Ethics*, *23*(3), 681–700.

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

Goodall, N. J. (2014). Ethical decision making during automated vehicle crashes. *Transportation Research Record*, *2424*(1), 58–65.

Gourlay, M. J., & Held, R. T. (2017). Head-mounted-display tracking for augmented and virtual reality [Wiley Online Library]. *Information Display*, *33*(1), 6–10.

Grafsgaard, J. F., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2012). Analyzing posture and affect in task-oriented tutoring. *Twenty-Fifth International FLAIRS Conference*.

Green, M. (2000). "how long does it take to stop?" methodological analysis of driver perception-brake times. *Transportation Human Factors*, *2*(3), 195–216.

Greger, R. (2018). Haftungsfragen beim automatisierten fahren. zum arbeitskreis II des verkehrsgerichtstags 2018. *NZV-Neue Zeitschrift fuer Verkehrsrecht*, *31*(1).

Gruyer, D., Magnier, V., Hamdi, K., Claussmann, L., Orfila, O., & Rakotonirainy, A. (2017). Perception, information processing and modeling: Critical stages for autonomous driving applications. *Annual Reviews in Control*, *44*, 323–341.

Grzywaczewski, A. (2017, October 10). *Training AI for self-driving vehicles: The challenge of scale* [NVIDIA developer blog]. Retrieved August 21, 2021, from https://developer.nvidia.com/blog/training-self-driving-vehicles-challenge-scale/

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, *51*(5), 1–42.

Gunes, H., & Pantic, M. (2010). Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners. *International conference on intelligent virtual agents*, 371–377.

Gurney, J. K. (2013). *Sue my car not me: Products liability and accidents involving autonomous vehicles* [HeinOnline]. Retrieved August 9, 2021, from https://heinonline.org/HOL/LandingPage?handle=hein.journals/jltp2013&div=15&id=&page=

Haber, R. N. (1978). Visual perception. *Annual review of psychology*, *29*(1), 31–59.

Haghighi, M. R. R., Sayari, M., Ghahramani, S., & Lankarani, K. B. (2020). Social, economic, and legislative factors and global road traffic fatalities. *BMC Public Health*, *20*(1), 1413.

Hagman, O. (2010). Driving pleasure: A key concept in swedish car culture. *Mobilities*, *5*(1), 25–39.

Hallerbach, S., Eberle, U., & Koester, F. (2017). The challenges of releasing cooperative and highly automated vehicles - a look beyond functional requirements. *AmE 2017 - Automotive meets Electronics; 8th GMM-Symposium*, 1–5.

Hammal, Z., Cohn, J. F., Heike, C., & Speltz, M. L. (2015). What can head and facial movements convey about positive and negative affect? *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, 281–287.

Hamon, R., Junklewitz, H., & Sanchez, I. (2020). Robustness and explainability of artificial intelligence. *Publications Office of the European Union*.

Hars, A. (2016, November 30). *Transformations 2025: How volkswagen prepares for the (driverless?) future | driverless car market watch* [Driverless-Future.com]. Retrieved May 28, 2021, from https://www.driverless-future.com/?p=1019

Hevelke, A., & Nida-Rümelin, J. (2015a). Ethische fragen zum verhalten selbstfahrender autos bei unausweichlichen unfällen: Der schutz von unbeteiligten. *Zeitschrift für philosophische Forschung*, *69*(2), 217–224. Retrieved June 1, 2021, from https://www.jstor.org/stable/24587552

Hevelke, A., & Nida-Rümelin, J. (2015b). Responsibility for crashes of autonomous vehicles: An ethical analysis. *Science and Engineering Ethics*, *21*(3), 619–630.

Hevelke, A., & Nida-Rümelin, J. (2015c). Selbstfahrende Autos und Trolley-Probleme: Zum Aufrechnen von Menschenleben im Falle unausweichlicher Unfälle [De Gruyter Section: Jahrbuch für Wissenschaft und Ethik]. *Jahrbuch für Wissenschaft und Ethik*, *19*(1), 5–24.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, *349*(6245), 261–266.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(3), 407–434.

Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: What it is and how to measure it. *Proceedings of the symposium on eye tracking research and applications*, 45–52.

Horwitz, J. (2021, October 1). *The facebook files - a wall street journal investigation* [The Wall Street Journal]. Retrieved November 7, 2021, from https://www.wsj.com/articles/the-facebook-files-11631713039

Howard, D., & Dai, D. (2014). Public perceptions of self-driving cars: The case of berkeley, california. *Transportation research board 93rd annual meeting 4502*, *14*, 1–16.

Hua, H., Krishnaswamy, P., & Rolland, J. P. (2006). Video-based eyetracking methods and algorithms in head-mounted displays [Optical Society of America]. *Optics Express*, *14*(10), 4328–4350.

Hussain, R., & Zeadally, S. (2019). Autonomous cars: Research results, issues, and future challenges. *IEEE Communications Surveys & Tutorials*, *21*(2), 1275–1313.

Hwang, Y. H. (2015). IoT security & privacy: Threats and challenges. *Proceedings of the 1st ACM Workshop on IoT Privacy, Trust, and Security*, 1–1.

Interrante, V., Hollerer, T., & Lecuyer, A. (2018). Virtual and augmented reality. *IEEE Computer Graphics and Applications*, *38*(2), 28–30.

IT Commitee of the United Nations. (1968). *Final act of the united nations conference on road traffic*. Retrieved August 9, 2021, from https://treaties.un.org/doc/Treaties/1977/05/19770524%5C%2000-13%5C%20AM/Ch_XI_B_19.pdf

Jarosch, O., Bellem, H., & Bengler, K. (2019). Effects of task-induced fatigue in prolonged conditional automated driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *61*(7), 1186–1199.

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

Johnson, T. (2013). Enhancing safety through automation. *Society of Automotive Engineers Gov't Industry Meeting, Automation and Connected Vehicle Safety, NHTSA*.

Johnson-Laird, P. N. (1986). Mental models: Towards a cognitive science of language, interface, and consciousness [Harvard University Press]. *Cognitive Science*, *6*.

Johnson-Laird, P. N. (2006). *How we reason*. Oxford University Press.

Juliani, A., Berges, V.-P., Teng, E., Cohen, A., Harper, J., Elion, C., Goy, C., Gao, Y., Henry, H., Mattar, M., & Lange, D. (2020). Unity: A general platform for intelligent agents. *arXiv preprint:1809.02627*. Retrieved June 10, 2021, from http://arxiv.org/abs/1809.02627

Kallioinen, N., Pershina, M., Zeiser, J., Nosrat Nezami, F., Pipa, G., Stephan, A., & König, P. (2019). Moral judgements on the actions of self-driving cars and human drivers in dilemma situations from different perspectives. *Frontiers in Psychology*, *10*, 2415.

Kalloc Studios, Inc. (2013). *Fuzor 3d virtual construction design*. Retrieved June 10, 2021, from https://www.kalloctech.com/

Karwowski, W. (2006). *International encyclopedia of ergonomics and human factors, -3 volume set*. Crc Press.

Keshava, A., Aumeistere, A., Izdebski, K., & Konig, P. (2020). Decoding task from oculomotor behavior in virtual reality. *ACM Symposium on Eye Tracking Research and Applications*, 1–5.

Keshava, A., Nezami, F. N., Neumann, H., Izdebski, K., Schueler, T., & Koenig, P. (2021). Just-in-time: Gaze guidance behavior while action planning and execution in VR. *bioRxiv preprint*. https://www.biorxiv.org/content/10.1101/2021.01.29.428782v1?rss=1

Kim, H. J., & Yang, J. H. (2015). Design of LDWS based on performance-based approach considering driver behaviors. *Journal of Institute of Control, Robotics and Systems*, *21*(11), 1081–1087.

Kim, J., & Canny, J. (2017). Interpretable learning for self-driving cars by visualizing causal attention. *Proceedings of the IEEE international conference on computer vision*, 2942–2950.

Knuth, K. (2013). Optimal data-based binning for histograms, 30, doi: Abs. *arXiv preprint physics/0605197*, *605*.

König, M., & Neumayr, L. (2017). Users' resistance towards radical innovations: The case of the self-driving car [Elsevier BV]. *Transp. Res. Part F Traffic Psychol. Behav.*, *44*, 42–52.

König, S. U., Clay, V., Nolte, D., Duesberg, L., Kuske, N., & König, P. (2019). Learning of spatial properties of a large-scale virtual city with an interactive map. *Front. Hum. Neurosci.*, *13*, 240.

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2015). Why did my car just do that? explaining semi-autonomous driving actions to improve driver understanding, trust, and performance [Springer Science and Business Media LLC]. *Int. J. Interact. Des. Manuf.*, *9*(4), 269–275.

Koo, J., Shin, D., Steinert, M., & Leifer, L. (2016). Understanding driver responses to voice alerts of autonomous car operations [Inderscience Publishers]. *Int. J. Veh. Des.*, *70*(4), 377.

Kozuka, S. (2019). A governance framework for the development and use of artificial intelligence: Lessons from the comparison of japanese and european initiatives. *Uniform Law Review*, *24*(2), 315–329.

Kröger, F. (2015). Das automatisierte Fahren im gesellschaftsgeschichtlichen und kulturwissenschaftlichen Kontext. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.), *Autonomes Fahren* (pp. 41–67). Springer Berlin Heidelberg.

Krueger, R., Rashidi, T. H., & Rose, J. M. (2016). Preferences for shared autonomous vehicles [Elsevier BV]. *Transp. Res. Part C: Emerg. Technol.*, *69*, 343–355.

Kuss, M., Jäkel, F., & Wichmann, F. A. (2005). Bayesian inference for psychometric functions. *Journal of Vision*, *5*(5), 8.

Kutila, M., Pyykonen, P., Ritter, W., Sawade, O., & Schaufele, B. (2016). Automotive LIDAR sensor development scenarios for harsh weather conditions. *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 265–270.

Kyriakidis, M., Happee, R., & Winter, J. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transp. Res. Part F Traffic Psychol. Behav.*, *32*.

Lague, S. (2021, June 8). *SebLague/path-creator* [Github repository]. Retrieved June 10, 2021, from https://github.com/SebLague/Path-Creator

Land, M. F. (2004). The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental Brain Research*, *159*(2), 151–160.

LaViola, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, *32*(1), 47–56.

Lee, C., & Coughlin, J. F. (2015). PERSPECTIVE: Older adults' adoption of technology: An integrated approach to identifying determinants and barriers [Wiley]. *J. Prod. Innov. Manage.*, *32*(5), 747–759.

Lee, C., Seppelt, B., Reimer, B., Mehler, B., & Coughlin, J. F. (2019). Acceptance of vehicle automation: Effects of demographic traits, technology experience and media exposure. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 2066–2070.

Lee, C., Ward, C., Raue, M., D'Ambrosio, L., & Coughlin, J. F. (2017). Age differences in acceptance of self-driving cars: A survey of perceptions and attitudes. *Human Aspects of IT for the Aged Population. Aging, Design and User Experience*, 3–13.

Lee, J.-G., Kim, K. J., Lee, S., & Shin, D.-H. (2015). Can autonomous vehicles be safe and trustworthy? effects of appearance and autonomy of unmanned driving systems. *International Journal of Human-Computer Interaction*, *31*(10), 682–691.

Lee, J.-H., & Song, C.-H. (2013). Effects of trust and perceived risk on user acceptance of a new technology service. *Social Behavior and Personality: an international journal*, *41*(4), 587–597.

Lee, J. D., & Kolodge, K. (2020). Exploring trust in self-driving vehicles through text analysis. *Human factors*, *62*(2), 260–277.

Lewis, J. R. (2018). The system usability scale: Past, present, and future. *International Journal of Human–Computer Interaction*, *34*(7), 577–590.

Li, A., Montaño, Z., Chen, V. J., & Gold, J. I. (2011). Virtual reality and pain management: Current trends and future directions. *Pain Management*, *1*(2), 147–157.

Li, H., Li, N., Kolmanovsky, I., & Girard, A. (2020). Energy-efficient autonomous vehicle control using reinforcement learning and interactive traffic simulations. *2020 American Control Conference (ACC)*, 3029–3034.

Li, J., Zhao, X., Cho, M.-J., Ju, W., & Malle, B. F. (2016). *From trolley to autonomous vehicle: Perceptions of responsibility and moral norms in traffic accidents with self-driving cars* (SAE Technical Paper No. 2016-01-0164). SAE International. Warrendale, PA.

Li, S., Blythe, P., Edwards, S., Goodman, P., & Hill, G. (2019). Investigation of the influence of multitasking on drivers' takeover performance in highly automated vehicles. *26th Intelligent Transport Systems World Congress*.

Lin, P. (2013, October 8). *The ethics of autonomous cars* [The atlantic]. Retrieved May 28, 2021, from https://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/

Lin, P. (2015). Why Ethics Matters for Autonomous Cars. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.), *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte* (pp. 69–85). Springer.

Lindgren, T., Fors, V., Pink, S., & Osz, K. (2020). Anticipatory experience in everyday autonomous driving. *Personal and Ubiquitous Computing*, *24*(6), 747–762.

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, *16*(3), 31–57.

Lohmann, M. F. (2016). Liability issues concerning self-driving vehicles. *European Journal of Risk Regulation*, *7*(2), 335–340.

Loyola-Gonzalez, O. (2019). Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE Access*, *7*, 154096–154113.

Lu, Y., & Sarter, N. (2019). Eye tracking: A process-oriented method for inferring trust in automation as a function of priming and system reliability. *IEEE Transactions on Human-Machine Systems*, *49*(6), 560–568.

Lüders, M., & Brandtzæg, P. B. (2017). 'my children tell me it's so simple': A mixed-methods approach to understand older non-users' perceptions of social networking sites. *New Media & Society*, *19*(2), 181–198.

Lutz, L. S. (2020). Haftung am beispiel automatisierter fahrzeuge. *KI & recht kompakt* (pp. 117–142). Springer.

Madary, M., & Metzinger, T. K. (2016). Real virtuality: A code of ethical conduct. recommendations for good scientific practice and the consumers of VR-technology. *Frontiers in Robotics and AI*, *3*.

Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., & Cusimano, C. (2015). Sacrifice one for the good of many? people apply different moral norms to human and robot agents. *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 117–124.

Marberger, C., Mielenz, H., Naujoks, F., Radlmayr, J., Bengler, K., & Wandtner, B. (2018). Understanding and applying the concept of "driver availability" in automated driving. In N. A. Stanton (Ed.), *Advances in human aspects of transportation* (pp. 595–605). Springer International Publishing.

Martinez, R. (2019). Artificial intelligence: Distinguishing between types & definitions. *Nevada Law Journal*, *19*(3), 9.

Masalonis, A. J., Duley, J. A., & Parasuraman, R. (1999). Effects of manual and autopilot control on mental workload and vigilance during simulated general aviation flight. *Transportation Human Factors*, *1*(2), 187–200.

Maurer, M. (2015). Einleitung. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.), *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte* (pp. 1–8). Springer.

McBride, N. (2016). The ethics of driverless cars. *ACM SIGCAS Computers and Society*, *45*(3), 179–184.

McCarthy, J. (2007). What is artificial intelligence?

McLachlan, G. J. (2005). *Discriminant analysis and statistical pattern recognition*. John Wiley & Sons.

Mehrabian, A. (2017). Communication without words. *Communication theory* (pp. 193–200). Routledge.

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

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

Michon, J. A. (1985). A critical view of driver behavior models: What do we know, what should we do? *Human behavior and traffic safety* (pp. 485–524). Springer.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, *11*(4), 143–152.

Millar, J. (2016). An ethics evaluation tool for automating ethical decision-making in robots and self-driving cars. *Applied Artificial Intelligence*, *30*(8), 787–809.

Miller, A. (2018). The effect of virtual reality education tools on the retention of information. *South Carolina Junior Academy of Science*. https://scholarexchange.furman.edu/scjas/2018/all/199

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, *267*, 1–38.

Monett, D., Lewis, C. W. P., Thórisson, K. R., Bach, J., Baldassarre, G., Granato, G., Berkeley, I. S. N., Chollet, F., Crosby, M., Shevlin, H., Fox, J., Laird, J. E., Legg, S., Lindes, P., Mikolov, T., Rapaport, W. J., Rojas, R., Rosa, M., Stone, P., ... Winfield, A. (2020). Special issue "on defining artificial intelligence"—commentaries and author's response. *Journal of Artificial General Intelligence*, *11*(2), 1–100.

Morra, L., Lamberti, F., Prattico, F. G., Rosa, S. L., & Montuschi, P. (2019). Building trust in autonomous vehicles: Role of virtual reality driving simulators in HMI design. *IEEE Transactions on Vehicular Technology*, *68*(10), 9438–9450.

Morris, D. Z. (2016, October 15). *Mercedes-benz's self-driving cars would choose passenger lives over bystanders*. Fortune Magazine. Retrieved November 11, 2021, from http://fortune.com/2016/10/15/mercedes-self-driving-car-ethics/

Munn, S. M., & Pelz, J. B. (2008). 3d point-of-regard, position and head orientation from a portable monocular video-based eye tracker. *Proceedings of the 2008 symposium on Eye tracking research & applications*, 181–188.

Murray, C. J. (1994). Quantifying the burden of disease: The technical basis for disability-adjusted life years. *Bulletin of the World health Organization*, *72*(3), 429.

Mwaffo, V., Curry, J. S., Iudice, F. L., & De Lellis, P. (2020). Pause-and-go self-balancing formation control of autonomous vehicles using vision and ultrasound sensors. *IEEE Transactions on Control Systems Technology*, 1–13.

Nastjuk, I., Herrenkind, B., Marrone, M., Brendel, A. B., & Kolbe, L. M. (2020). What drives the acceptance of autonomous driving? an investigation of acceptance factors from an end-user's perspective [Elsevier]. *Technological Forecasting and Social Change*, *161*, 120319.

Naujoks, F., Forster, Y., Wiedemann, K., & Neukum, A. (2016). Speech improves human-automation cooperation in automated driving. *Mensch und Computer 2016–Workshopband*.

Navarrete, C. D., McDonald, M. M., Mott, M. L., & Asher, B. (2012). Virtual morality: Emotion and action in a simulated three-dimensional "trolley problem". *Emotion*, *12*(2), 364–370.

Nees, M. A. (2016). Acceptance of self-driving cars: An examination of idealized versus realistic portrayals with a self-driving car acceptance scale. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*, 1449–1453.

Newen, A., Bruin, L. D., & Gallagher, S. (Eds.). (2018, September). *The oxford handbook of 4e cognition*. Oxford University Press.

Nezami, F. N., Wächter, M. A., Maleki, N., Spaniol, P., Kühne, L. M., Haas, A., Pingel, J. M., Tiemann, L., Nienhaus, F., Keller, L., König, S. U., König, P., & Pipa, G. (2021). Westdrive x LoopAR: An open-access virtual reality project in unity for evaluating user interaction methods during takeover requests. *Sensors*, *21*(5), 1879.

Nezami, F. N., Wächter, M. A., Pipa, G., & König, P. (2020). Project westdrive: Unity city with self-driving cars and pedestrians for virtual reality studies. *Frontiers in ICT*, *7*, 1.

Nguyen, T. C. (2019, June 30). *History of self-driving cars - how a 1930's exhibit set the stage for driverless cars* [thoughtco.com]. Retrieved August 25, 2021, from https://www.thoughtco.com/history-of-self-driving-cars-4117191

Noë, A. (2009). *Out of our heads: Why you are not your brain, and other lessons from the biology of consciousness*. Macmillan.

Norman, D. A. (1990). The 'problem'with automation: Inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *327*(1241), 585–593.

Norman, D. A. (2015). The human side of automation. In G. Meyer & S. Beiker (Eds.), *Road vehicle automation 2* (pp. 73–79). Springer International Publishing. Retrieved October 14, 2021, from http://link.springer.com/10.1007/978-3-319-19078-5_7

Norris, W. R., & Patterson, A. E. (2019). System-level testing and evaluation plan for field robots: A tutorial with test course layouts. *Robotics*, *8*(4), 83.

Norton, P. (2008). Fighting traffic. *The Dawn of the Motor Age in the American City, Cambridge*.

Official Journal of the European Union. (2021, March 9). *UN regulation no 157 – uniform provisions concerning the approval of vehicles with regards to automated lane keeping systems [2021/389]OJL82/75* [Eur-lex.europa.eu] [Doc ID: 42021X0389]. Retrieved November 9, 2021, from https://eur-lex.europa.eu/eli/reg/2021/389/oj

Ogdon, D. C. (2019). HoloLens and VIVE pro: Virtual reality headsets. *Journal of the Medical Library Association*, *107*(1).

O'Meara, V. (2019). Weapons of the chic: *Instagram* influencer engagement pods as practices of resistance to *Instagram* platform labor. *Social Media + Society*, *5*(4), 205630511987967.

Ongsulee, P. (2017). Artificial intelligence, machine learning and deep learning. *2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)*, 1–6.

OpenStreetMap Foundation. (2004). *OpenStreetMap Collaborative mapping*. Retrieved June 10, 2021, from https://www.openstreetmap.org/

Othman, K. (2021). Public acceptance and perception of autonomous vehicles: A comprehensive review. *AI and Ethics*, *63*.

Ouyang, Z., Niu, J., Liu, Y., & Guizani, M. (2020). Deep CNN-based real-time traffic light detector for self-driving vehicles. *IEEE Transactions on Mobile Computing*, *19*(2), 300–313.

Pan, S., Fulton, L. M., Roy, A., Jung, J., Choi, Y., & Gao, H. O. (2021). Shared use of electric autonomous vehicles: Air quality and health impacts of future mobility in the united states. *Renewable and Sustainable Energy Reviews*, *149*, 111380.

Pan, X., Banakou, D., & Slater, M. (2011). Computer based video and virtual environments in the study of the role of emotions in moral behavior. In S. D'Mello, A. Graesser, B. Schuller, & J.-C. Martin (Eds.), *Affective computing and intelligent interaction* (pp. 52–61). Springer.

Pan, X., & Hamilton, A. F. d. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, *109*(3), 395–417.

Pan, Z., Cheok, A. D., Yang, H., Zhu, J., & Shi, J. (2006). Virtual reality and mixed reality for virtual learning environments. *Computers & graphics*, *30*(1), 20–28.

Papadoulis, A., Quddus, M., & Imprialou, M. (2019). Evaluating the safety impact of connected and autonomous vehicles on motorways [Elsevier]. *Accident Analysis & Prevention*, *124*, 12–22.

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in human neuroscience*, *9*, 660.

Paschek, D., Luminosu, C. T., & Draghici, A. (2017). Automated business process management – in times of digital transformation using machine learning or artificial intelligence. *MATEC Web of Conferences*, *121*, 04007.

Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuroscience*, *9*(1), 94–107.

Pek, J., Wong, O., & Wong, A. C. M. (2018). How to address non-normality: A taxonomy of approaches, reviewed, and illustrated. *Front. Psychol.*, *9*, 2104.

Penmetsa, P., Sheinidashtegol, P., Musaev, A., Adanu, E. K., & Hudnall, M. (2021). Effects of the autonomous vehicle crashes on public perception of the technology. *IATSS Research*.

Pohl, J., Sethsson, M., Degerman, P., & Larsson, J. (2006). A semi-automated parallel parking system for passenger cars. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, *220*(1), 53–65.

Prautzsch, H., Boehm, W., & Paluszny, M. (2002). *Bézier and b-spline techniques* (Vol. 6). Springer.

Proudlock, F. A., Shekhar, H., & Gottlob, I. (2003). Coordination of eye and head movements during reading. *Investigative Opthalmology & Visual Science*, *44*(7), 2991.

Puertas-Ramirez, D., Serrano-Mamolar, A., Martin Gomez, D., & Boticario, J. G. (2021). Should conditional self-driving cars consider the state of the human inside the vehicle? *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 137–141.

Raats, K., Fors, V., & Pink, S. (2020). Trusting autonomous vehicles: An interdisciplinary approach. *Transportation Research Interdisciplinary Perspectives*, *7*, 100201.

Ramos, S., Gehrig, S., Pinggera, P., Franke, U., & Rother, C. (2017). Detecting unexpected obstacles for self-driving cars: Fusing deep learning and geometric modeling. *2017 IEEE Intelligent Vehicles Symposium (IV)*, 1025–1032. https://doi.org/10.1109/IVS.2017.7995849

Rao, D., Pathrose, P., Huening, F., & Sid, J. (2019). An approach for validating safety of perception software in autonomous driving systems. *International Symposium on Model-Based Safety and Assessment*, 303–316.

Rasmussen, J. (1986). Information processing and human-machine interaction [Elsevier Science Publishing Company, Inc.]. *An approach to cognitive engineering*.

Raue, M., D'Ambrosio, L. A., Ward, C., Lee, C., Jacquillat, C., & Coughlin, J. F. (2019). The influence of feelings while driving regular cars on the perception and acceptance of self-driving cars [Wiley]. *Risk Anal.*, *39*(2), 358–374.

Reggente, N., Essoe, J. K.-Y., Aghajan, Z. M., Tavakoli, A. V., McGuire, J. F., Suthana, N. A., & Rissman, J. (2018). Enhancing the ecological validity of fMRI memory research using virtual reality. *Frontiers in Neuroscience*, *12*, 408.

Reimer, B. (2014). Driver assistance systems and the transition to automated vehicles: A path to increase older adult safety and mobility? *Public Policy & Aging Report*, *24*(1), 27–31.

Rezaei, A., & Caulfield, B. (2020). Examining public acceptance of autonomous mobility. *Travel Behaviour and Society*, *21*, 235–246.

Riva, G. (2003). Virtual environments in clinical psychology. *Psychotherapy: Theory, Research, Practice, Training*, *40*(1), 68.

Riva, G. (2005). Virtual reality in psychotherapy: Review. *CyberPsychology & Behavior*, *8*(3), 220–230.

Roche, F., Somieski, A., & Brandenburg, S. (2019). Behavioral changes to repeated takeovers in highly automated driving: Effects of the takeover-request design and the nondriving-related task modality. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *61*(5), 839–849.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215.

Rudin, C., & Wagstaff, K. L. (2014). Machine learning for science and society. *Machine Learning*, *95*(1), 1–9.

Rus-Calafell, M., Garety, P., Sason, E., Craig, T. J. K., & Valmaggia, L. R. (2018). Virtual reality in the assessment and treatment of psychosis: A systematic review of its utility, acceptability and effectiveness. *Psychological Medicine*, *48*(3), 362–391.

Ryan, M. (2020). The future of transportation: Ethical, legal, social and economic impacts of self-driving vehicles in the year 2025. *Sci. Eng. Ethics*, *26*(3), 1185–1208.

Sachdeva, S., Iliev, R., Ekhtiari, H., & Dehghani, M. (2015). The role of self-sacrifice in moral dilemmas. *PLOS ONE*, *10*(6), e0127409.

SAE Internation. (2014). J3016: Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems - SAE international. Retrieved December 3, 2021, from https://www.sae.org/standards/content/j3016_201401/

Schoettle, B. (2017). Sensor fusion: A comparison of sensing capabilities of human drivers and highly automated vehicles. *University of Michigan*.

Schreurs, M. A., & Steuwer, S. D. (2015). *Autonomous driving - political, legal, social, and sustainability dimensions* (M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner, Eds.). Springer Berlin Heidelberg.

Seppelt, B. D., & Lee, J. D. (2019). Keeping the driver in the loop: Dynamic feedback to support appropriate use of imperfect vehicle control automation [Elsevier]. *International Journal of Human-Computer Studies*, *125*, 66–80.

Shadrin, S. S., & Ivanova, A. A. (2019). Analytical reviw of SAE j3016: Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. *Avtomobil'. Doroga. Infrastruktura.*, (3), 10.

Shalev-Shwartz, S., Shammah, S., & Shashua, A. (2018). On a formal model of safe and scalable self-driving cars. *arXiv:1708.06374*. Retrieved August 30, 2021, from http://arxiv.org/abs/1708.06374

Shapiro, A. (2017). Reform predictive policing. *Nature news*, *541*(7638), 458.

Shaw, J., Rudzicz, F., Jamieson, T., & Goldfarb, A. (2019). Artificial intelligence and the implementation challenge. *Journal of Medical Internet Research*, *21*(7), e13659.

Shinq-Jen Wu, Hsin-Han Chiang, Jau-Woei Perng, Tsu-Tian Lee, & Chao-Jung Chen. (2005). The automated lane-keeping design for an intelligent vehicle. *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.*, 508–513.

Sikkenk, M., & Terken, J. (2015). Rules of conduct for autonomous vehicles. *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 19–22.

Silva, D., Földes, D., & Csiszár, C. (2021). Autonomous vehicle use and urban space transformation: A scenario building and analysing method [Multidisciplinary Digital Publishing Institute]. *Sustainability*, *13*(6), 3008.

Skulmowski, A., Bunge, A., Kaspar, K., & Pipa, G. (2014). Forced-choice decision-making in modified trolley dilemma situations: A virtual reality and eye tracking study. *Frontiers in Behavioral Neuroscience*, *8*.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *6*(6), 603–616.

Sloane, M., & Moss, E. (2019). AI's social sciences deficit. *Nature Machine Intelligence*, *1*(8), 330–331.

Sony Interactive Entertainment LLC. (2021). *PlayStation VR | live the game in incredible virtual reality worlds* [PlayStation.com]. Retrieved October 18, 2021, from https://www.playstation.com/en-us/ps-vr/

Spring, P., McIntosh, A., Caponecchia, C., Baysari, M., et al. (2009). Level of automation: Effects on train driver vigilance. *Proceedings of Third International Conference on Rail Human Factors, Lille, France, 3rd–5th March*.

Stanton, N. A., & Salmon, P. M. (2009). Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. *Safety Science*, *47*(2), 227–237.

Stephan, A. (2019). Trust in highly automated driving.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence [Wiley Online Library]. *Journal of communication*, *42*(4), 73–93.

Summala, H. (2000). Brake reaction times and driver behavior analysis. *Transportation Human Factors*, *2*(3), 217–226.

Sütfeld, L. R., Gast, R., König, P., & Pipa, G. (2017). Using virtual reality to assess ethical decisions in road traffic scenarios: Applicability of value-of-life-based models and influences of time pressure. *Frontiers in Behavioral Neuroscience*, *11*, 122.

Swaroop, D., & Hedrick, J. K. (1999). Constant spacing strategies for platooning in automated highway systems. *Journal of Dynamic Systems, Measurement, and Control*, *121*(3), 462–470.

Szikora, P., & Madarász, N. (2017). Self-driving cars — the human side. *2017 IEEE 14th International Scientific Conference on Informatics*, 383–387.

Takacs, A., Rudas, I., Bosl, D., & Haidegger, T. (2018). Highly automated vehicles and self-driving cars [industry tutorial]. *IEEE Robotics & Automation Magazine*, *25*(4), 106–112.

Tesla Motors. (2020). *Tesla Autopilot*. Retrieved June 9, 2021, from https://www.tesla.com/autopilot

Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, *59*(2), 204–217.

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*(6), 1395.

Tian, J., Xie, H., Hu, S., & Liu, J. (2021). Multidimensional face representation in a deep convolutional neural network reveals the mechanism underlying AI racism. *Frontiers in Computational Neuroscience*, *15*, 620281.

Tulshan, A. S., & Dhage, S. N. (2018). Survey on virtual assistant: Google assistant, siri, cortana, alexa. *International symposium on signal processing and intelligent recognition systems*, 190–201.

Unger, P. K. (1996). *Living high and letting die: Our illusion of innocence*. Oxford University Press.

Unity Technologies. (2021). *Unity Echtzeit-Entwicklungsplattform | 3D-, 2D-VR- & AR-Engine* [Unity.com]. Retrieved October 15, 2021, from https://unity.com/de

Valve Corporation. (2021). *Steamvr a suite of technologies created by valve that power high-fidelity VR experiences.* Retrieved October 18, 2021, from https://www.steamvr.com/en/

van Emmerik, M. L., de Vries, S. C., & Bos, J. E. (2011). Internal and external fields of view affect cybersickness [Elsevier]. *Displays*, *32*(4), 169–174.

van Tongern, S. (2018). *Hub chain mobility*. Retrieved October 14, 2021, from https://www.hubchain.de/

Varghese, J. Z., Boone, R. G. et al. (2015). Overview of autonomous vehicle sensors and systems. *International Conference on Operations Excellence and Service Engineering*, 178–191.

Veale, M. (2020). A critical take on the policy recommendations of the EU high-level expert group on artificial intelligence. *European Journal of Risk Regulation*, 1–10. https://doi.org/10.1017/err.2019.65

Veitengruber, J., Boucek, G., & Smith, W. (1977). *Aircraft alerting systems criteria study* (Vol. 1).

Veličković, P., & Milovanović, M. (2021). Improvement of the interaction model aimed to reduce the negative effects of cybersickness in VR rehab applications. *Sensors*, *21*(2), 321.

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, *11*(4), 342–365.

Verma, S. (2019). Weapons of math destruction: How big data increases inequality and threatens democracy. *Vikalpa: The Journal for Decision Makers*, *44*(2), 97–98.

Wachenfeld, W., Winner, H., Gerdes, C., Lenz, B., Maurer, M., Beiker, S. A., Fraedrich, E., & Winkle, T. (2015). Use-Cases des autonomen Fahrens. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.), *Autonomes Fahren* (pp. 9–37). Springer Berlin Heidelberg.

Waldrop, M. M. (2019). News feature: What are the limits of deep learning? *Proceedings of the National Academy of Sciences*, *116*(4), 1074–1077.

Ward, M. R., Lee, C., D'Ambrosio, L., F, J., & M. (2017). Coughlin,"acceptance of automated driving across generations: The role of risk and benefit perception, knowledge, and trust. *HCI 2017, lecture notes in ComputerScience, vol.10271* (pp. 254–266). Springer.

Warne, R. T. (2014). A primer on multivariate analysis of variance (MANOVA) for behavioral scientists [Citeseer]. *Practical Assessment, Research & Evaluation*, *19*.

Warwick, K. (2012). *Artificial intelligence: The basics*. Routledge.

Waschl, H., Schmied, R., Reischl, D., & Stolz, M. (2019). A virtual development and evaluation framework for ADAS—case study of a p-ACC in a connected environment. *Control strategies for advanced driver assistance systems and autonomous driving functions* (pp. 107–131). Springer.

Wattanakriengkrai, S., Chinthanet, B., Hata, H., Kula, R. G., Treude, C., Guo, J., & Matsumoto, K. (2020). GitHub repositories with links to academic papers: Open access, traceability, and evolution. *arXiv preprint :2004.00199*. https://arxiv.org/abs/2004.00199

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117.

Wedlin, J., Tillback, L.-R., & Bane, O. (1992). Combining properties for driving pleasure and driving safety: A challenge for the chassis engineer, 921595.

Weller, G., & Schlag, B. (2007). Driver behaviour model: Ripcord-inserest deliverable d8. *Technische Universität Dresden, Dresden*.

Werner, A. (2020). Organizing music, organizing gender: Algorithmic culture and spotify recommendations. *Popular Communication*, *18*(1), 78–90.

Weyer, J., Fink, R. D., & Adelt, F. (2015). Human–machine cooperation in smart cars. an empirical investigation of the loss-of-control thesis. *Safety Science*, *72*, 199–208.

Wienrich, C., Schindler, K., Dollinger, N., Kock, S., & Traupe, O. (2018). Social presence and cooperation in large-scale multi-user virtual reality - the relevance of social interdependence for location-based environments. *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 207–214.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636.

Winfield, A. F. T., Blum, C., & Liu, W. (2014). Towards an ethical robot: Internal models, consequences and ethical action selection. In M. Mistry, A. Leonardis, M. Witkowski, & C. Melhuish (Eds.), *Advances in autonomous robotics systems* (pp. 85–96). Springer International Publishing.

Wintersberger, P., Frison, A.-K., Riener, A., & Sawitzky, T. v. (2019). Fostering user acceptance and trust in fully automated vehicles: Evaluating the potential of augmented reality. *PRESENCE: Virtual and Augmented Reality*, *27*(1), 46–62.

Wintersberger, P., Nicklas, H., Martlbauer, T., Hammer, S., & Riener, A. (2020). Explainable automation: Personalized and adaptive uis to foster trust and understanding of driving automation systems. *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 252–261.

Wintersberger, P., & Riener, A. (2016). Trust in technology as a safety aspect in highly automated driving. *i-com*, *15*(3), 297–310.

Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The dark sides of artificial intelligence: An integrated AI governance framework for public administration. *International Journal of Public Administration*, *43*(9), 818–829.

Wolfartsberger, J. (2019). Analyzing the potential of virtual reality for engineering design review. *Automation in Construction*, *104*, 27–37.

World Health Organization. (2021, June 21). *Road traffic injuries - fact sheet*. Retrieved August 25, 2021, from https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Yadav, A. K., & Velaga, N. R. (2020). Alcohol-impaired driving in rural and urban road environments: Effect on speeding behaviour and crash probabilities. *Accident Analysis & Prevention*, *140*, 105512.

Yang, J., & Coughlin, J. F. (2014). In-vehicle technology for self-driving cars: Advantages and challenges for aging drivers. *Int. J. Automot. Technol. Manage.*, *15*(2), 333–340.

Yang, M., Wang, S., Bakita, J., Vu, T., Smith, F. D., Anderson, J. H., & Frahm, J.-M. (2019). Re-thinking CNN frameworks for time-sensitive autonomous-driving applications: Addressing an industrial challenge. *2019 IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)*, 305–317.

Yaqoob, I., Khan, L. U., Kazmi, S. A., Imran, M., Guizani, N., & Hong, C. S. (2019). Autonomous driving cars in smart cities: Recent advances, requirements, and challenges. *IEEE Network*, *34*(1), 174–181.

Yarbus, A. L. (1967). Eye movements during perception of complex objects. *Eye movements and vision* (pp. 171–211). Springer.

Yoo, H. W., Druml, N., Brunner, D., Schwarzl, C., Thurner, T., Hennecke, M., & Schitter, G. (2018). MEMS-based lidar for autonomous driving. *e & i Elektrotechnik und Informationstechnik*, *135*(6), 408–415.

Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, *2*(10), 719–731.

Yuan, Y. (2021, January 22). *Base station setting* [Pimax support]. Retrieved November 18, 2021, from https://support.pimax.com/en/support/solutions/articles/60000691051-base-station-setting

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

Zhang, B., de Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *64*, 285–307.

Zhao, Y., Wang, X., Goubran, M., Whalen, T., & Petriu, E. M. (2013). Human emotion and cognition recognition from body language of the head using soft computing techniques. *Journal of Ambient Intelligence and Humanized Computing*, *4*(1), 121–140.