TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM SCHOOL OF ENGINEERING AND DESIGN

Comparison of Methods to Evaluate Automated Vehicle's Driving Behavior for Communicating Intentions to Pedestrians by Using Wizard of Oz, Virtual Reality, and Video Setups

Tanja Elisabeth Fuest

Vollständiger Abdruck der von der TUM School of Engineering and Design
der Technischen Universität München
zur Erlangung des akademischen Grades
Doktorin der Ingenieurswissenschaften (Dr.-Ing.)
genehmigten Dissertation.

Vorsitz:		Prof. Dr.-Ing. Karsten Stahl
Prüfer der Dissertation:	1.	Prof. Dr. Klaus Bengler
	2.	Prof. Dr. Mark Vollrath

Die Dissertation wurde am 22.04.2021 bei der Technischen Universität München
eingereicht und durch die Fakultät für Maschinenwesen am 28.09.2021 angenommen.

For my parents, who made me the person I am today.

Acknowledgement

This thesis could not have been completed without the support of numerous people. I would like to take this opportunity to thank all of them for their help.

First of all, I would like to thank my supervisor **Prof. Klaus Bengler** for encouraging me in my ideas, guiding me with suggestions and supporting me in difficult situations. In addition, he made it possible for me to network beyond the European borders, from which I benefited greatly.

I would also like to thank **Prof. Mark Vollrath** for his engagement at the doctoral workshops and the congress of the DGP. This enables helpful exchange among like-minded people.

This thesis was part of a cooperative project with BMW. During the most intensive phases of the project, it was especially **Dr. Elisabeth Schmidt** who provided me with helpful advice.

I would like to thank **Dr. Thomas Streubel** for taking over the mentorship. I enjoyed our scientific exchange at different meetings and conferences.

Moreover, my thanks goes to my students who have supported me over the past years. In particular, I would like to thank **Annika Kirner** and **Lars Michalowski** for their contributions to my thesis.

Furthermore, I would like to emphasize the good teamwork between the colleagues at the Chair of Ergonomics. I enjoyed every constructive coffee session as well as the scientific and private exchange during and after work, especially with **André**, **Deike**, **Luis**, **Oliver** and **Tobias**.

Last but not least, I am more than grateful for the encouragement of my family and friends. **Katrin** improved my English by proofreading my publications and this thesis. My parents **Hildegard** and **Norbert** and my brother **Hendrik** have supported and motivated me throughout my entire life. Last but not least, my biggest thanks goes to **Florian**, who helped me through the entire time with his expertise in macros and programming, but also with good coffee, patience and his constant faith in me.

Abstract

The integration of automated vehicles into traffic poses a number of challenges. One of these challenges is designing the communication of automated vehicles with human road users. In this regard, it is essential to design the communication in a way that is understandable for all human road users and ensures an efficient and safe interaction. In recent years, there have been many studies focusing on the design of driving behavior and additional external human machine interfaces, like displays or projections, especially with regard to communication between automated vehicles and pedestrians. These studies used different setups, e.g., videos, virtual reality, or Wizard of Oz approaches. A comparison of the study results showed that the chosen method seems to have an impact on the outcome of the study. However, the exact extent of transferability of the results between these studies is still unknown. Therefore, the aim of the thesis is to give recommendations which method should be used to evaluate the automated vehicle's driving behavior to communicate a yielding or non-yielding intention to a pedestrian.

A taxonomy of traffic situations was created to systematically investigate the influence of the method (Article 1). The taxonomy includes all attributes that could influence the interaction between automated vehicles and other human road users. In addition, a metric—the Intention Recognition Time—and a questionnaire were developed to differentiate between different driving profiles. The metric and questionnaire items allow for comparing study setups without exposing participants to the risk of an accident with the automated vehicle. The metric, as well as the situation derived from the taxonomy, were first evaluated in a virtual reality setup (Article 2). The study was followed by an evaluation of the metrics and situation in a Wizard of Oz setup (Article 3). Another Wizard of Oz study tested the extent to which a driver can replicate a given target driving profile multiple times (Article 4). Subsequently, this Wizard of Oz study was supplemented by a virtual reality and two video studies for a systematic comparison of methods (Article 5).

The results showed that it is possible to differentiate between unambiguous and ambiguous driving profiles in all setups. This means that unsuitable driving profiles can be detected with all setups. However, the results of the virtual reality and video setups differed in some aspects from the results of the Wizard of Oz setup. This can be explained in particular by perceptual differences in virtual reality and video use. Therefore, the Wizard of Oz setup is recommended for evaluating different driving profiles for automated vehicles to communicate with pedestrians.

Zusammenfassung

Die Integration von automatisierten Fahrzeugen in den Straßenverkehr birgt einige Herausforderungen. Eine davon ist die Gestaltung der Kommunikation mit anderen Verkehrsteilnehmern. Dazu gehört, dass die Kommunikation von automatisierten Fahrzeugen so gestaltet werden muss, dass sie für alle Verkehrsteilnehmer verständlich ist und ein effizientes und sicheres Miteinander gewährleistet. In den letzten Jahren wurden viele Studien durchgeführt, die sich mit der Gestaltung von Fahrprofilen und zusätzlichen Displays oder Projektionen als Mittel zur Kommunikation auseinandergesetzt haben, insbesondere im Hinblick auf die Kommunikation zwischen automatisierten Fahrzeugen und Fußgängern. Diese Studien verwendeten verschiedene Methoden, zum Beispiel Videos, Virtual Reality oder Wizard of Oz. Ein Vergleich der Studienergebnisse zeigt, dass das gewählte Setup einen Einfluss auf das Ergebnis der Studie haben kann. Das genaue Ausmaß der Übertragbarkeit der Ergebnisse zwischen diesen Studien ist jedoch noch unbekannt. Das Ziel der Dissertation ist es daher, Empfehlungen abzuleiten, mit welcher Methode das Fahrverhalten eines automatisierten Fahrzeugs evaluiert werden sollte. Mit dem Fahrverhalten soll dabei eindeutig kommuniziert werden, dass ein Fußgänger vorgelassen oder nicht vorgelassen wird.

Um den Einfluss der Methode systematisch zu untersuchen, wurde im Rahmen der Dissertation zunächst eine Taxonomie von Verkehrssituationen erstellt, die alle Attribute darstellt, die die Interaktion zwischen automatisierten Fahrzeugen und anderen Straßenverkehrsteilnehmern beeinflussen könnten (Artikel 1). Zusätzlich wurden eine Metrik, die Intentionserkennungszeit (Intention Recognition Time), und ein Fragebogen entwickelt, mit denen zwischen verschiedenen Fahrprofilen differenziert werden soll. Diese Metriken ermöglichen einen Vergleich der Setups, ohne die Teilnehmer dem Risiko eines Unfalls mit dem automatisierten Fahrzeug auszusetzen. Die Metriken sowie die aus der Taxonomie abgeleitete Situation wurden zunächst in einer Virtual Reality Studie evaluiert (Artikel 2). Im Anschluss erfolgte die Überprüfung der Metriken und der Situation in einer Wizard of Oz Studie (Artikel 3). In einer weiteren Wizard of Oz Studie wurde geprüft, inwieweit ein Fahrer ein vorgegebenes Fahrprofil replizieren kann (Artikel 4). Im Anschluss wurde diese Wizard of Oz Studie durch eine Virtual Reality und zwei Video Studien ergänzt, um einen systematischen Methodenvergleich durchzuführen (Artikel 5).

Die Ergebnisse zeigen, dass es bei Nutzung der Metriken in allen verwendeten Methoden möglich ist, zwischen eindeutigen und uneindeutigen Fahrprofilen zu differenzieren. Dies bedeutet, dass ungeeignete Fahrprofile mit allen Methoden detektiert werden. Allerdings weichen die Ergebnisse der Virtual Reality und Video Studien in einigen Punkten von den Ergebnissen der Wizard of Oz Methode ab. Dies lässt sich insbesondere durch Wahrnehmungsunterschiede bei der Nutzung von Virtual Reality und Videos erklären. Daher wird für die Evaluierung von verschiedenen Fahrprofilen für automatisierte Fahrzeuge zur Kommunikation mit Fußgängern die Wizard of Oz Methode empfohlen.

Table of Contents

1 Introduction 1

2 Communication in Road Traffic 3

- 2.1 Explicit Communication ... 5
  - 2.1.1 Positive Impact of Explicit Communication ... 6
  - 2.1.2 Negative or No Impact of Explicit Communication ... 7
- 2.2 Implicit Communication ... 8
- 2.3 Summary of Communication in Road Traffic ... 10

3 Methods 11

- 3.1 Image ... 11
- 3.2 Video ... 12
- 3.3 Virtual Reality ... 13
- 3.4 Driving Simulator ... 13
- 3.5 Wizard of Oz ... 14
- 3.6 Experimental Vehicle ... 15
- 3.7 Summary of Methods ... 16

4 Metrics 19

- 4.1 Objective Data ... 19
- 4.2 Subjective Data ... 21
- 4.3 Summary of Metrics ... 22

5 Derivation of Metrics 23

6 Objectives and Contributions 25

7 Taxonomy of Traffic Situations for the Interaction between Automated Vehicles and Human Road Users 26

8 How Should an Automated Vehicle Communicate Its Intention to a Pedestrian? – A Virtual Reality Study 27

9 Using the Driving Behavior of an Automated Vehicle to Communicate Intentions – A Wizard of Oz Study 28

10 Reproducibility of Driving Profiles – Application of the Wizard of Oz Method for Vehicle Pedestrian Interaction 29

11 Comparison of Methods to Evaluate the Influence of Automated Vehicle's Driving Behavior on Pedestrians: Wizard of Oz, Virtual Reality, and Video 30

12 General Discussion 31

- 12.1 Comparison of the Methods ... 31
  - 12.1.1 Wizard of Oz ... 31
  - 12.1.2 Virtual Reality ... 32
  - 12.1.3 Videos ... 33
  - 12.1.4 Limitations of the Comparison ... 33
- 12.2 Recommendations for Metrics to Evaluate the Driving Behavior ... 36
  - 12.2.1 Objective Metrics ... 36
  - 12.2.2 Subjective Metrics ... 37
  - 12.2.3 Transferability of the Metrics ... 39
- 12.3 Recommendations for the Driving Profile of Automated Vehicles ... 40
  - 12.3.1 Driving Profile for the Intention "AV goes first" ... 40
  - 12.3.2 Driving Profile for the Intention "Let the HRU go first" ... 41
  - 12.3.3 Transferability of the Driving Profiles ... 43

13 Future Work 45

14 Conclusion 47

15 Complementary Studies and Contributor Roles 48

16 References 51

List of Figures

Figure 2.1 Sender Receiver Mode 3
Figure 2.2 Explicit communication 5
Figure 3.1 Wizard of Oz approaches 15
Figure 4.1 The Model Human Processor 19
Figure 12.1 Vehicles used as AVs in the different setups 35
Figure 12.2 Pictures from the video setups 35
Figure 12.3 Target driving profiles 37

List of Tables

Table 3.1 Advantages and Disadvantages of Methods 17
Table 7.1 Overview of chosen attributes and value facets for the studies of the thesis 26
Table 12.1 Driving profiles for the intention “AV goes first” 41
Table 12.2 Driving profiles for the intention “Let the HRU go first” 42

Abbreviations

AV	Automated Vehicle
CAVE	Cave Automatic Virtual Environment
eHMI	external Human Machine Interface
HRU	Human Road User
IRT	Intention Recognition Time
LED	Light Emitting Diode
SAE	Society of Automotive Engineers
StVO	Straßenverkehrs–Ordnung
TTA	Time to Arrival
VR	Virtual Reality
WoZ	Wizard of Oz

1 Introduction

The first vehicles with partially automated systems (Society of Automotive Engineers (SAE) level 2; SAE, 2018) are already driving on highways (Cacilo et al., 2015). For example, a traffic jam assistant is able to take over the longitudinal and lateral control of the vehicle in low speed ranges on highways (Cacilo et al., 2015; Robert Bosch GmbH, 2020). However, the driver has to monitor the system permanently and has to be able to take over the driving task at any time (Cacilo et al., 2015; Robert Bosch GmbH, 2020). In the long term, highly automated systems (SAE level 4; SAE, 2018) intend to completely disengage the driver from the driving task. These systems should be able to maneuver the vehicle without the driver's supervision on highways and later on also in urban traffic. Urban traffic is more complex than highway traffic (Schneemann & Gohl, 2016). The complexity arises from ambiguous scenarios (Clausen & Klingner, 2018), a high number of static and dynamic objects, and different types of human road users (HRUs) (Hipp et al., 2018), such as pedestrians, cyclists (Lüke et al., 2015), and human drivers, taking part in the traffic system. Therefore, before integrating automated vehicles (AVs) in urban traffic some questions still need to be clarified. This includes questions about technical implementations, the communication with the passenger—e.g., to inform about ongoing driving maneuvers—, and the communication of AV's with HRUs. The latter has been investigated in particular for drivers (e.g., Rettenmaier et al., 2019; Rettenmaier, Albers, & Bengler, 2020) and pedestrians (e.g., Ackermann et al., 2018; Ackermann, Beggiato, Bluhm et al., 2019; Moore et al., 2019). This is particularly relevant for pedestrians as they are considered the most unpredictable and vulnerable road users (Deb, Rahman et al., 2018; Rasouli & Tsotsos, 2018). A deadly accident in 2018 between a self-driving Uber and a pedestrian in Arizona (United States) (Griggs & Wakabayashi, 2018) was probably caused by a chain of unfortunate circumstances. The pedestrian was not correctly classified due to a software error (Shepardson, 2020). In addition, the emergency brake assistant was deactivated and the safety driver, as back-up in case of system failure, was probably distracted by using a mobile phone (Shepardson, 2020). This emphasizes the need of enhancing the technology and the algorithms to correctly detect and classify the HRU, as well as the need to keep the driver in the loop in partially automated systems. Moreover, it illustrates that it is advisable to inform pedestrians about the AV's intention. Even if this type of communication would not have been possible in the example due to the software error, there are other situations in which an unambiguous communication of the AV might help to avoid accidents or deadlock situations.

While the German road traffic regulations (Straßenverkehrs–Ordnung, StVO; Bundesamt für Justiz, 2013) contain rules for most situations, there are still some situations that have to be solved using communication between the road users involved. Examples for such situations are bottlenecks (Imbsweiler et al., 2017; Rettenmaier et al., 2019) and shared space situations, such as parking areas (Fuest, Maier et al., 2020; Witzlack et al., 2016). These situations are characterized by only few legal regulations (Witzlack et al., 2016), so that communication between road users is unavoidable. With

AVs, however, a new communication partner enters the traffic system. In order for AVs to move smoothly and safely, the implemented algorithms must not only be able to recognize potential sources of danger, but also learn to understand existing communication forms in traffic. In addition, AVs have to learn to communicate with other HRUs to avoid accidents. AVs can use existing communication, e.g., predesigned driving profiles, or add external human-machine interfaces (eHMIs) to communicate intentions (chapter 2).

A lot of research was conducted within the last years to evaluate the use of driving behavior and eHMIs to communicate intentions to HRUs. Methods used to evaluate communication are, e.g., images (e.g., Fridman et al., 2019, Yang, 2017), videos (e.g., Ackermann et al., 2018, Petzoldt et al., 2018), virtual reality (VR) (e.g., Burns et al., 2020, Deb, Strawderman, & Carruth, 2018), Wizard of Oz (WoZ) (e.g., Moore et al., 2019, Rothenbücher et al., 2016), and experimental vehicles (e.g., Potzy, Feinauer et al., 2019) (chapter 3). Besides the methods, different metrics are used to evaluate communication, such as the crossing time (e.g., Deb, Strawderman, & Carruth, 2018; Dietrich et al., 2018), error or misinterpretation rate (e.g., Chang et al., 2017; Weber et al., 2019), or the gap acceptance (e.g., Maruhn et al., 2020; Schneider et al., 2021) (chapter 4). For example, when comparing different driving profiles or eHMIs, the driving profile or eHMI that results in an earlier crossing time and/or a lower error or misinterpretation rate is the most beneficial. In addition, the gap acceptance can be used to evaluate what gap sizes pedestrians accept to cross the street. Both—the method and the chosen metric—may influence the conclusion of the findings. For example, pedestrians might indicate that they accept a defined gap size, but if they were asked to actually cross in front of the vehicle, they might not feel safe enough.

This thesis aims to compare different methods—video, VR, and WoZ—to investigate which method can be used to evaluate how an AV should drive to communicate a yielding or a non-yielding intention to a pedestrian. Metrics are needed for this comparison which can be applied to all methods (chapter 5). The objectives of the thesis are explained in more detail in chapter 6.

To answer the research questions, a taxonomy of traffic situations was created to systematically investigate the influence of the method (chapter 7). The metric as well as the situation derived from the taxonomy were evaluated in a VR setup (chapter 8) and a WoZ setup (chapter 9). In another WoZ study the extent to which a driver can replicate a given target driving profile multiple times was investigated (chapter 10). This research was followed by the systematic comparison of methods: two different video setups, a VR, and a WoZ setup (chapter 11).

The results from all studies regarding the comparison of methods and recommendations for metrics and driving profiles are discussed in relation to the associated research questions (chapter 12). Subsequently, recommendations for future work are given (chapter 13) and a conclusion (chapter 14) is drawn.

2 Communication in Road Traffic

Our everyday life is characterized by communication and interaction with other humans to convey our needs and to achieve our goals (Nimmons, 2014). Communication always follows a similar pattern as described by the *Sender Receiver Model* (Shannon, 1948; Shannon & Weaver, 1998): a sender wants to transmit a message to at least one receiver (Figure 2.1). The sender transmits the message by encoding the message, e.g., in spoken words. Subsequently, the receiver has to decode the message (Shannon & Weaver, 1998). Transmission errors, so-called noise, may occur during the transmission of the message: for example, distance, culture, and language differences could interfere with the transmission (Chandler & Munday, 2011). A common criticism of the model is that communication is described as a one-way process. However, usually the communication partner reacts to the message. Hence, the model was adapted to include feedback from the receiver to the sender (Figure 2.1; Wiener, 2000).

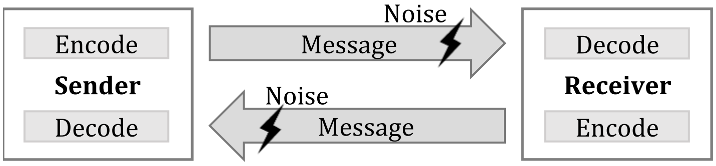

Figure 2.1 Sender Receiver Model based on Shannon and Weaver (1998) and Wiener (2000).

The model has been designed for technical purposes, in particular to improve telephone communication (Röhner & Schütz, 2016). The model deals with the transmission and reception of messages, but not with the meaning of the message itself. Other models include such a subconscious message (e.g., the *Four-Sides Model* by Schulz von Thun, 2018, or the *Five Axioms of Communication* by Watzlawick, 2016). It can be assumed that subliminal messages are also communicated in road traffic, as traffic behavior can be understood as a social behavior with interactions between road users (Klebelsberg, 1982). Although, if an AV is a technical systems, it has already been shown that humans are polite to technical systems (Reeves & Nass, 1997) and anthropomorphize vehicles (Waytz et al., 2014). Nevertheless, it is important that an AV is able to encode a message so that a HRU can decode the message correctly. Therefore, the communication design needs to be basic so that a noise-free communication can be ensured before a social component is taken into account.

Therefore, the *Sender Receiver Model* is transferred to the communication in road traffic in a first step. Road users need communication to ensure that traffic can move forward safely and efficiently. However, messages from drivers are not transmitted via spoken words but with light or sound signals, i.e., explicit communication, as well as with the driving behavior, i.e., implicit communication (Fuest, Sorokin et al., 2018). Just as in the *Sender Receiver Model* the messages can be disturbed by noise during transmission. Distance, culture, and language differences can lead to errors in communication not only in linguistic usage, but also during communication in road traffic. It is therefore important to evaluate the communication of AVs across countries (Weber et al., 2019) and to have precisely defined communication regulations. This is helpful for AVs to decode the message of HRUs and to be able to encode their own messages.

For Germany, the formal traffic regulations can be found in the StVO (Bundesamt für Justiz, 2013). However, the StVO leaves room for interpretation. For example, the StVO stipulates the timely use of the indicators when the driver wants to turn right (StVO, §9; Bundesamt für Justiz, 2013). It also requires drivers to indicate the intention by driving as far as possible on the right side of the lane (StVO, §9; Bundesamt für Justiz, 2013). However, it is not further specified when or at what distance to the junction the driver should do so. Another example is the correct behavior when a pedestrian is standing at a crosswalk. Here, pedestrians have priority over vehicles, except for railway vehicles (StVO, §26; Bundesamt für Justiz, 2013). It is stipulated that drivers must approach the crosswalk at moderate speed and brake when pedestrians indicate that they wish to use the crosswalk (StVO, §26; Bundesamt für Justiz, 2013). However, it is not regulated what moderate speed means and how pedestrians have to indicate their crossing intention.

AVs have to learn how to handle this room of interpretation in the StVO correctly. For example, if two pedestrians stand facing away from the road near a crosswalk and talk to each other, an experienced driver recognizes through the implicit signals of the pedestrians' body language, that they have no crossing intention or that they will yield for the vehicle. As the experiences from the Bertha Benz drive show, an AV is not necessarily able to assess this situation correctly and would slow down initially (Daimler AG, 2020). However, pedestrians wanting to cross the road often wait for signals that the driver is going to yield. For example, eye contact or hand gestures are used to communicate yielding intentions (Rouchitsas & Alm, 2019). This informal communication occurs wherever a formal action has not been specified, such as in parking areas (Witzlack et al., 2016).

The examples of missing informal communication show that the communication with AVs will differ from the communication with drivers (Habibovic et al., 2018). Therefore, before integrating AVs in mixed traffic—with pedestrians, cyclists, drivers, and other AVs—they have to be able to apply formal regulations and to understand informal communication to decode the messages from other HRUs. Therefore, the question arises, whether the informal communication, such as eye contact or hand gestures (Färber, 2016), should be replaced as this communication is not achievable for AVs. When it needs to be replaced, it has to be specified by what means and to what extent (Bengler, Rettenmaier et al., 2020).

Replacing informal communication can be achieved by using explicit and implicit communication (Breazeal et al., 2005). Thus, besides using predesigned driving profiles for implicit communication (e.g., Burns et al., 2020; Potzy, Feinauer et al., 2019), it is also possible for AVs to use built-in technology to install Light Emitting Diode (LED) strips (e.g., Weber et al., 2019), displays (e.g., Rettenmaier, Albers, & Bengler, 2020) or projections (e.g., Dietrich et al., 2018) by which the AV can explicitly communicate (Figure 2.2). These two communication options—explicit and implicit communication—, however, must be designed and evaluated for AVs in advance (Fuest, Sorokin et al., 2018).


(a) **(b)** **(c)**

Figure 2.2 Explicit communication: **(a)** LED strip (Weber et al., 2019, p. 517); **(b)** display (Rettenmaier, Albers, & Bengler, 2020, p. 179); **(c)** projection (Dietrich et al., 2018, p. 46).

2.1 Explicit Communication

Communicating explicitly is defined as transmitting content via direct signs (Gildert et al., 2018). Explicit communication is used in spoken language, e.g., if drivers say "I'm going to turn right" they explicitly communicate the intention to a passenger. However, it can also be related to body language, as specific gestures transmit an explicit meaning (Gildert et al., 2018), such as "thumbs up" meaning "great" in Germany (Wagner, 2014). Transferring explicit communication onto road traffic, drivers can give pedestrians a direct sign to cross the road by waving the hand. Besides this informal communication, also formal explicit communication exists: using the right indicator is an explicit way to communicate the intention to turn right. Other examples are activating the hazard warning lights to inform about a traffic jam, or using the horn to communicate a direct warning.

AVs can use the existing forms of explicit communication, such as lights and sounds. Furthermore, AVs have the possibility to communicate via eHMIs. Such eHMIs may consist of LED strips (e.g., Faas et al., 2020; Faas & Baumann, 2020; Fuest, Feierle et al., 2020; Hensch et al., 2020b; Weber et al., 2019; Zhang et al., 2018), displays (e.g., Clercq et al., 2019; Joisten et al., 2020; Rettenmaier et al., 2019; Rettenmaier, Albers, & Bengler, 2020; Song et al., 2018), or projections (e.g., Dietrich et al., 2018; Kühn et al., 2019). In addition, acoustic signals can be used to communicate with HRUs (e.g., Deb, Strawderman, & Carruth, 2018).

While some eHMIs only inform the surrounding traffic about the status of the vehicle (automated vs. manual) (e.g., Fuest, Feierle et al., 2020; Kühn et al., 2019), other eHMIs communicate the AV's intention to pedestrians (e.g., Clercq et al., 2019; Dietrich et al., 2018; Eisma et al., 2020; Faas et al., 2020; Habibovic et al., 2018; Petzoldt et al., 2018; Weber et al., 2019) or drivers (e.g., Rettenmaier et al., 2019; Rettenmaier, Albers, & Bengler, 2020). However, in all cases the AV should only communicate its intention, instead of recommending actions to other HRUs (Andersson et al., 2017; Zhang et al., 2018).

The evaluation of eHMIs in studies concluded a positive effect of some eHMIs on the crossing behavior of pedestrians (e.g., Clercq et al., 2019; Eisma et al.; Petzoldt et al., 2018). However, other studies showed that not all participants understand the intentions transmitted by the eHMIs immediately (e.g., Clercq et al., 2019; Habibovic et al., 2018;

Hensch et al., 2020a). In the following sections, the advantages and disadvantages of explicit communication are presented.

2.1.1 Positive Impact of Explicit Communication

External communication is recommended to replace the existing communication between drivers and pedestrians (Hudson et al., 2019). Study results indicated that pedestrians crossed the road more often (Song et al., 2018), sooner (Othersen et al., 2019), and quicker (Chang et al., 2017) when an eHMI was used. Furthermore, when using eHMIs pedestrians crossed the road before the vehicle stopped completely (Habibovic et al., 2018; Othersen et al., 2019). This has the benefit that the AV does not decelerate to a standstill which saves fuel. In addition, the avoidance of a standstill allows a higher average speed, making the AV drive more efficiently. Deadlock situations in which both road users came to a complete standstill could be reduced by 38% when presenting an eHMI (Matthews et al., 2017). In addition, pedestrians understood the intention of an AV earlier, when a frontal brake light was presented (Petzoldt et al., 2017, 2018). They felt safer seeing an eHMI compared with a baseline condition, when the AV intended to yield (Clercq et al., 2019). Besides the perceived safety, also the comfort increased with an eHMI (Böckle et al., 2017). Pedestrians felt to be more in control of the situation and calmer when encountering an AV using an eHMI (Habibovic et al., 2018). In addition, the eHMI improved the correct intention recognition rate for participants from Germany and the United States for a yielding AV, when a light band was used (Weber et al., 2019). Therefore, it can be concluded that eHMIs improve the pedestrian–AV interaction (Deb, Strawderman, & Carruth, 2018) and increase trust (Matthews et al., 2017). The latter can be increased in particular if pedestrians have prior knowledge about the eHMIs and the vehicle (Matthews et al., 2017). Moreover, the confidence increased with longer interactions with AVs (Miguel et al., 2019).

The eHMI should be installed on the roof, windscreen, or the grille of the AV to increase the pedestrians' perceived safety to cross the road (Eisma et al., 2020). Even if projections or eHMIs on wheels also increase the feeling of being able to cross the road safely, they should be avoided as they are not in the direct field of view of pedestrians (Eisma et al., 2020). Moreover, the eHMI is understood immediately (Clercq et al., 2019), better (Chang et al., 2018), and is preferred (Deb, Strawderman, & Carruth, 2018) when it contains text.

Besides pedestrians, drivers are also able to understand explicit communication used by an AV. For a lane change, it is crucial to use the indicator to communicate the intention of merging (Potzy, Feuerbach, & Bengler, 2019). An early use of the indicator to inform others about a lane change is recommended, as it is perceived more cooperative by other drivers (Kauffmann, Winkler et al., 2018). Using eHMIs in a bottleneck situation results in a shorter passing time (Rettenmaier et al., 2019; Rettenmaier, Albers, & Bengler, 2020) with fewer crashes (Rettenmaier, Albers, & Bengler, 2020).

2.1.2 Negative or No Impact of Explicit Communication

Besides the positive effects of eHMIs, some studies revealed contradictory results and derived less clear or no recommendations for using eHMIs. A study result indicates that pedestrians are more likely to stick to their familiar behavior than to use the information of an eHMI (Clamann et al., 2017). Furthermore, depending on the location, eHMIs remain unnoticed, for example if they are installed behind the AV windshield (Miguel et al., 2019). Especially for non-yielding AVs no effect of eHMIs could be found (Clercq et al., 2019; Weber et al., 2019). Pedestrians did not feel safer with an eHMI (Clercq et al., 2019) and did not understand the AV's non-yielding intention earlier (Weber et al., 2019), as compared to a baseline without eHMI. In addition, the crossing initiation of pedestrians was not influenced by the presence of an eHMI (Dietrich, Boos et al., 2020). Moreover, an eHMI that increased the understanding of an AV's yielding intention in Germany and the United States, does not necessarily have an advantage in China (Weber et al., 2019).

Furthermore, pedestrians have to learn the meaning of most eHMIs, such as front brake lights, smileys, animated LED bands, and symbols, as they are not intuitive (Ackermann, Beggiato, Bluhm et al., 2019; Ackermann, Beggiato, Schubert, & Krems, 2019; Clercq et al., 2019; Deb, Strawderman, & Carruth, 2018; Habibovic et al., 2018; Hensch et al., 2020a; Mahadevan et al., 2018). Therefore, intuitive eHMIs are recommended to to gain an early understanding of the eHMI by pedestrians and to minimize learning effects (Rettenmaier et al., 2019). In addition, LED light bands are not advised because pedestrians did not relate the communication with LED light bands to themselves (Hensch et al., 2020a).

There are different opinions as to what should be depicted on eHMIs. While Clercq et al. (2019) recommended eHMIs with text, other researchers pointed out that light and beep signals were preferred to eHMIs with text (Merat et al., 2018). Moreover, Löcken et al. (2019) argued that using text depends on the size of the display. This is in line with the results from Rettenmaier, Schulze, and Bengler (2020), who point out that text has to be larger than symbols to maintain legibility. In addition, Chang et al. (2018) found a positive impact on eHMIs with virtual eyes, whereas Löcken et al. (2019) criticized that pedestrians felt more unsafe and did not trust systems with virtual eyes. They concluded that virtual eyes may be better than no eHMI, but inferior to other eHMI concepts (Löcken et al., 2019). Comparing two eHMIs that communicate the vehicle‘s maneuvering intentions by using either something like an indicator, or projecting the AVs path onto the ground revealed a positive effect on pedestrians hypothetical crossing behavior for the projection (Burns et al., 2019). However, both eHMIs were not free of criticism, as trust, acceptance and perceived safety did not differ between both eHMIs and pedestrians had to look at the ground and not at the vehicle when deciding to cross the road (Burns et al., 2019). The method of projection especially reaches its limits in bad or sunny weather conditions (Burns et al., 2019).

It has been shown that eHMIs as features of the automated system reach a higher trust score by pedestrians than the AV as a whole. Thus, an eHMI seems to be interpreted separately from the AV by pedestrians. However, the effect is reversed in case of malfunctions: pedestrians perceived the whole AV as faulty and not only the display if the eHMI transmits incorrect information. If information was transmitted incorrectly by an erroneous eHMI pedestrians' trust, perceived safety and confidence decreased. However, these attributes were quickly recovering. This effect can be seen as an indication of overtrust, which is defined as a misjudgment of risk when interacting with a system. (Holländer, Wintersberger, & Butz, 2019)

Furthermore, when evaluating interaction between AVs and drivers negative impacts can be found for malfunctions. The subsequent change of intention, as a result of a sensor error, caused longer passing times, decreased acceptance and reduced perceived usefulness in a bottleneck scenario (Rettenmaier, Albers, & Bengler, 2020). Therefore, an eHMI design should avoid misleading information (Holländer, Wintersberger, & Butz, 2019) and a flawless communication for AVs is required (Rettenmaier, Albers, & Bengler, 2020). In addition, a training for HRUs is recommended to better understand the functions and messages of eHMIs, also with regard to possible malfunctions (Holländer, Wintersberger, & Butz, 2019).

Regarding a marking of AVs as automated, it was shown that the pedestrians' willingness to cross the road (Dey et al., 2019), the critical gap, and the perceived level of stress (Rodríguez Palmeiro, 2017) were independent from the driving status (manual vs. automated) and the type of marking. However, when pedestrians recognized an AV as a self-driven vehicle, the feeling of uncertainty increased, the perceived safety decreased and they were more careful (Rodríguez Palmeiro, 2017). In addition, cyclists have more confidence in manually driven vehicles than in AVs as they do not expect to be detected more often by an AV (Hagenzieker et al., 2020). Therefore, the marking might have negative effects on cyclists. Furthermore, no effect was found for the communication of the AV's driving status to other drivers on highways (Fuest, Feierle et al., 2020; Kühn et al., 2019). However, if an AV was marked incorrectly, drivers had difficulties attributing driving behavior to a human or AV (Kühn et al., 2019).

2.2 Implicit Communication

In implicit communication "the practical behavior itself is the message" (Castelfranchi et al., 2010, p. 2). This means that by performing actions, messages can be sent to a communication partner (Castelfranchi et al., 2010). To refer to the example in section 2.1, the driver could drive as far as possible on the right side of the lane and lose speed to communicate the intention of turning right. Even if the indicator is missing, the driver behind the vehicle would probably understand that the vehicle will turn right. Implicit communication therefore tends to convey intentions in an indirect way. Pedestrians (Dey & Terken, 2017; Rasouli et al., 2018b), as well as drivers (Dietrich et al., 2019) mostly use implicit communication. For example, drivers prefer implicit communication in bottlenecks to inform other drivers about their intentions (Imbsweiler et al., 2017). In addition, drivers using explicit communication to communicate with pedestrians can hardly be observed (Lee et al., 2020).

Moreover, drivers and pedestrians are able to understand implicit communication, even from AVs (Lundgren et al., 2017; Potzy, Feinauer et al., 2019; Potzy, Feuerbach, & Bengler, 2019). Therefore, AVs can use implicit communication, e.g., to merge on highways (Potzy, Feinauer et al., 2019), or to brake for pedestrians to tell them that the AV is yielding (e.g., Dey et al., 2019; Dietrich, Maruhn et al., 2020). Clamann et al. (2017) revealed that pedestrians rely on proven strategies, such as driving behavior, rather than on eHMIs. Even if research focused on the evaluation of eHMIs, pedestrians mentioned that their crossing decision was based on the AVs kinematics (Li et al., 2018).

Based on the motto "You cannot not communicate" (Watzlawick, 2016), the AV's driving behavior is, in contrast to eHMIs, constantly observable by HRUs as long as there is no visual impairment. Therefore, AVs have to learn to communicate by using implicit communication, such as the driving behavior (Lundgren et al., 2017). Conventional driving behavior plays a major role because results showed that pedestrians mistrust atypical driving profiles, as compared to conventional driving profiles (H. Schmidt et al., 2019).

Pedestrians use the driving behavior, e.g., speed, position of the vehicle, and the distance between the AV and the pedestrian (Burns et al., 2020; Li et al., 2018; Rasouli et al., 2018a; S. Schmidt & Färber, 2009) to decide whether they can cross the road. Additionally, feeling comfortable is connected to the distance of the AV (Matthews et al., 2017). An early deceleration is recommended to communicate an AV's yielding intention (Ackermann et al., 2018; Ackermann, Beggiato, Bluhm et al., 2019; Fuest, Maier et al., 2020; Fuest, Michalowski et al., 2018). Artificial pitches helped to reduce the time between a standstill of the AV and the moment the pedestrians enter the hypothetical encroachment zone; however, they were rated poorly and some pedestrians interpreted the pitch as an emergency maneuver or a defect on the AV (Dietrich, Maruhn et al., 2020). Pedestrians perceive an AV as very polite if it decelerates to a standstill, in this case all pedestrians crossed in front of the AV (Moore et al., 2019).

On highways with dense traffic, drivers expect an efficient lane change from AVs with a small time headway when merging in front of the driver (Potzy, Feinauer et al., 2019). This decreases the disturbance and increases the acceptance and the predictability of the AV's intention (Potzy, Feinauer et al., 2019). A strong deceleration to the target gap is suggested to inform other drivers of a lane change (Potzy, Feuerbach, & Bengler, 2019). The strong deceleration is, besides a speed reduction and fast reaction times, observed as most cooperative (Kauffmann, Naujoks et al., 2018). Additionally, a merging vehicle is perceived as more cooperative, when it uses an early longitudinal acceleration when beginning the lane change (Kauffmann, Winkler et al., 2018).

In particular, the published studies give recommendations for driving profiles. Negative aspects are seldom discussed in the papers. Therefore, a chapter with negative aspects of implicit communication is not included in the thesis.

2.3 Summary of Communication in Road Traffic

Both implicit and explicit communication can be used to communicate an AV's intention to pedestrians and drivers. It seems that pedestrians prefer proven methods, i.e., the driving behavior instead of eHMIs to decide whether they can cross a road (Clamann et al., 2017). Research also indicates that marking an AV or additional communication through eHMIs is not required to transmit communications to other HRUs because the driving behavior is sufficient (Moore et al., 2019). When using eHMIs, it must be ensured that malfunctions do not occur or at least that their influence will not lead to accidents. It is also important to ensure that eHMIs are understood by all HRUs. Previous study results indicate that most eHMIs are not intuitively understandable (Ackermann, Beggiato, Bluhm et al., 2019; Ackermann, Beggiato, Schubert, & Krems, 2019; Clercq et al., 2019; Habibovic et al., 2018; Hensch et al., 2020a). However, eHMIs could be useful as a secondary information channel (Li et al., 2018). In addition, results illustrate that it is possible to shorten HRUs' decision times through eHMIs.

Nevertheless, since the employed driving profile has not been published in all studies so far, it is possible that eHMIs only compensate for a poor driving profile. Besides these facts, researchers cannot ignore the importance of designing the AV's driving behavior, as it is constantly observable by HRUs in case there is no visual impairment. Therefore, before designing eHMIs, an AV's driving behavior for different traffic situations has to be evaluated. For this reason, the thesis uses the driving behavior when comparing different methods. As a result, recommendations for AV's driving behavior will be derived.

3 Methods

Several methods are used to design and evaluate AVs' explicit and implicit communication in order to communicate intentions to HRUs, particularly pedestrians or drivers. Therefore, the chapter focuses on the description of the advantages and disadvantages of methods (Table 3.1.) that are used to evaluate the AV's communication to pedestrians and drivers. Other methods, such as bicycle simulators, augmented reality, and cave automatic virtual environment (CAVE) studies are not considered, as there is little literature regarding the AV's communication available.

For each method, the three quality criteria should be applied: objectivity, reliability and validity. Especially for the WoZ method special requirements must be met in order to fulfill these criteria. A test is objective if the result is independent of the investigator (Bengler, 2015). Thus, in a WoZ setup, the results need to be independent of the driving wizard (Müller et al., 2019). Reliability describes the accuracy of a measuring instrument (Bengler, 2015). One possibility to determine the reliability is the retest method (Sedlmeier & Renkewitz, 2011). For the WoZ approach, this means that the wizard needs to reproduce the driving profile in every trial (Müller et al., 2019). Validity is given when a test measures the dimension that it intends to measure (Sedlmeier & Renkewitz, 2011). Transferred to the WoZ approach, this means, i.e., that participants believe they encounter a "real" AV (Müller et al., 2019). A distinction can be made between internal and external validity. External validity refers to the transferability of the results (Vollrath, 2015). Internal validity exist when the effect can be unequivocally attributed to the influence of the independent variable (Vollrath, 2015). Thus, the groups studied may differ only in the independent variable and confounding variables can be excluded (Vollrath, 2015). The quality criteria are interdependent: without objectivity, high reliability cannot be achieved, and reliability is the prerequisite for validity (Sedlmeier & Renkewitz, 2011). A validation of the method is not performed within the scope of this thesis. However, it is examined whether the participants perceive the simulated AV as real.

3.1 Image

Images without moving elements are an economical way to evaluate communication of AVs with pedestrians by collecting first impressions for early-stage design concepts (Fridman et al., 2019; Fuest, Schmidt, & Bengler, 2020). The concepts, such as altered eHMIs, can be designed and easily edited to different vehicle types by using computer software. Therefore, a large amount of eHMIs concepts (Fridman et al., 2019; Yang, 2017), marked AVs (Hagenzieker et al., 2020), and different vehicle types (Dey et al., 2019) can be evaluated without risk and within a short amount of time. In addition, different driver states, e.g., sleeping, reading, or paying attention, were pictured and participants assessed their willingness to cross the street (Lundgren et al., 2017). When using images in an online survey, a large group of participants worldwide can evaluate the different communication concepts at low cost (Rice et al., 2017; Wright, 2005). However, participants as well as external influences cannot be controlled and therefore the validity (Farmer, 1998) and reliability of online studies might be comparatively low. In addition, the perception of eHMIs via images may differ from the perception in reality. Particularly, as reliability in online studies is comparatively low if the conditions are not the same for all participants (e.g., different displays with dissimilar resolutions, and/or distraction by background noise).

Participants might be able to imagine themselves in different situations, when showing them images of these situations. This can be, e.g., used to evaluate pedestrians' crossing behavior by asking participants to imagine how they normally react in a situation with a specific vehicle's driving behavior (Fuest, Maier et al., 2020). However, their answers can differ from their real behavior, as images are not immersive. Therefore, the application is limited, as driving behavior cannot be visualized.

The evaluation of use cases, with surrounding traffic, distraction or visual coverings, is not possible using images. Due to this reason and the limitations of an image setup, a final validation of design concepts in another setup is needed before integrating such concepts in traffic.

3.2 Video

As compared to images, videos have the advantage that driving behavior can be visualized. However, it remains unclear to which extent the visualized driving behavior differs from the perception in real traffic (Petzoldt et al., 2018). Just like images, videos are cost-efficient, risk-free, less immersive (as compared to VR, WoZ or experimental vehicle setups), and can be made available to participants using online surveys.

A distinction can be made between different types of videos. While in some approaches the videos were recorded in traffic or on test tracks to ensure a realistic environmental setting (e.g., Ackermann et al., 2018; Ackermann, Beggiato, Schubert, & Krems, 2019; Beggiato et al., 2017; Dey et al., 2017; Fuest, Schmidt, & Bengler, 2020; Song et al., 2018), other approaches used videos recorded in VR (e.g., Fuest, Schmidt, & Bengler, 2020). In addition, it is possible to analyze videos of real traffic scenarios collected during a study (Cœugnet et al., 2019), through traffic monitoring, or on YouTube (Brown & Laurier, 2017) to gather information about the interaction between vehicles and passengers, as well as the interaction between vehicles and other drivers. There are also differences in how often participants are allowed to watch the video sequence. In some studies, participants were allowed to watch the videos as often as they wanted (Ackermann, Beggiato, Schubert, & Krems, 2019; Li et al., 2018) whereas in other studies the video was presented only once to the participants (Dey et al., 2017, 2019; Fuest, Schmidt, & Bengler, 2020).

Videos are already used to assess hypothetical crossing behavior of pedestrians depending on different AV's driving profiles (e.g., Ackermann et al., 2018; Ackermann, Beggiato, Bluhm et al., 2019; Dey et al., 2019; Fuest, Schmidt, & Bengler, 2020) or eHMIs (e.g., Ackermann, Beggiato, Schubert, & Krems, 2019; Chang et al., 2018; Eisma et al., 2020; Li et al., 2018; Petzoldt et al., 2018; Song et al., 2018).

3.3 Virtual Reality

A CAVE or a head-mounted display can be used for a VR setup. Visualizing a traffic situations in a VR setup is more immersive (Löcken et al., 2019), as compared to videos or images. In addition, pedestrians perceive the displayed environment as relatively realistic (Deb et al., 2017; Löcken et al., 2019). Results from Deb, Strawderman, and Carruth (2018) demonstrated that pedestrians react in VR as they would in real traffic when crossing a road. The pedestrians' walking speed matched their speed in real traffic (Deb et al., 2017). Moreover, it was shown that the perception of distance and subjective danger by a Segway in VR and real spaces did not differ (Iryo-Asano et al., 2018). However, there are also results that suggest that the perception in VR differs from real values (Bhagavathula et al., 2018; Schneider et al., 2021). The results from H. Schmidt et al. (2019) demonstrated that VR can be used to evaluate AV's driving behavior: pedestrians reacted with confusion and mistrust to atypical driving profiles, as compared to conventional driving profiles. It can be assumed that a VR simulator is a valid method to evaluate pedestrians' road crossing behavior (Deb et al., 2017).

In addition to the advantage that a realistic behavior can be reproduced in VR, it is risk-free (Burns et al., 2020; Deb et al., 2017) when evaluating high-risk situations because pedestrians never encounter a real vehicle. Another benefit is the controlled environment, which enables a high reliability for VR studies (Deb et al., 2017). Just like video or image setups, a study design can be implemented relatively cost-efficient (Burns et al., 2020; Deb et al., 2017). Only, a one-off investment in the required technology must be made. In addition, a short amount of time is needed to evaluate different driving profiles or eHMIs in altered trials (Deb et al., 2017). Depending on the study design, not much space is required for the VR setup and it can be easily transported and reinstalled elsewhere (Weber et al., 2019). However, the space requirement changes as large laboratories must be available if pedestrians are asked to cross a virtual road (Deb et al., 2017). Additionally, the movement in VR can lead to simulation sickness (Deb et al., 2017) and the perception in VR using a head-mounted display seems to deviate from real values for the estimation of speed (Bhagavathula et al., 2018).

Within the last years, an increasing number of VR studies have been conducted to evaluate the influence of AV's driving behavior (e.g., Burns et al., 2020; Dietrich, Maruhn et al., 2020; Fuest, Maier et al., 2020; Jayaraman et al., 2018; H. Schmidt et al., 2019) or eHMIs (e.g., Böckle et al., 2017; Chang et al., 2017; Clercq et al., 2019; Deb, Strawderman, & Carruth, 2018; Holländer, Colley et al., 2019; Holländer, Wintersberger, & Butz, 2019; Hudson et al., 2019; Weber et al., 2019) on pedestrians.

3.4 Driving Simulator

While VR studies focused on the communication between AVs and pedestrians, driving simulator studies were used to design the communication between AVs and drivers. Similar to VR, driving simulator studies are highly immersive; however, this depends on the simulator being static or dynamic (Kantowitz, 2011). In addition, also the investment in the technology (and if necessary in the license) has to be made. The perception of speed in the simulator varies from the perception in reality (Hurwitz et al., 2005). As compared to experimental vehicle studies, situations can be evaluated, just like in videos and VR, risk-free for participants, even if safety-critical situations are evaluated. However, participants need a familiarization phase, as the motion in the simulator differ from real vehicles (Greenberg & Blommer, 2011). As a result, simulator sickness might occur (Allen et al., 2011). As in VR studies, conditions can be kept constant and thus high reliability can be achieved.

Two scenarios have been investigated so far: lane changes (Kauffmann, Naujoks et al., 2018; Kauffmann, Winkler et al., 2018) and encounters in bottlenecks (Rettenmaier et al., 2019; Rettenmaier, Albers, & Bengler, 2020). For this purpose, drivers encounter an AV that merges lanes with different driving profiles (e.g., Kauffmann, Naujoks et al., 2018; Kauffmann, Winkler et al., 2018) or use eHMIs to communicate the AV’s intention (e.g., Rettenmaier et al., 2019; Rettenmaier, Albers, & Bengler, 2020). In addition, driving simulator studies were used to evaluate the influence of marking AVs on highways (e.g., Fuest, Feierle et al., 2020; Kühn et al., 2019).

3.5 Wizard of Oz

Before implementing explicit or implicit communication into AVs it seems useful to evaluate concepts and driving profiles in real traffic or on test tracks. Since AVs and underlying algorithms are under development AVs can only be used for research purposes to a limited extent (Bengler, Omozik, & Müller, 2020). A possible solution is a WoZ approach. The WoZ paradigm was already applied to investigate the human–computer interaction with not fully developed systems (Fraser & Gilbert, 1991). Examples are intelligent tutoring systems (Bengler, Omozik, & Müller, 2020), and speech systems (Fraser & Gilbert, 1991). In those approaches, the investigator was hidden from the users and simulated the system (Fraser & Gilbert, 1991). This enables parallel technical development and validation of the systems (Bengler, Omozik, & Müller, 2020).

In addition, WoZ vehicles are already used to investigate the interaction between driver assistance systems and the driver inside the vehicle (G. Schmidt et al., 2008). Rothenbücher et al. (2016) were some of the first researchers who used WoZ vehicles to simulate an AV and evaluated the reaction by pedestrians. Seat covers are used to hide the driver from the outside perspective (Figure 3.1 (a); Currano et al., 2018; Fuest, Michalowski et al., 2018; Fuest, Michalowski et al., 2019; Fuest, Schmidt, & Bengler, 2020; Joisten et al., 2020; Moore et al., 2019; Rothenbücher et al., 2016). Other WoZ approaches used a right-hand steered vehicle (Figure 3.1 (b); Habibovic et al., 2016; Habibovic et al., 2018; Lagström & Lundgren, 2015), or a joystick, which was controlled by the co-driver (Rodríguez Palmeiro, 2017). Results demonstrated that pedestrians do not notice the driver under the seat cover (Dietrich, Boos et al., 2020; Rothenbücher et al., 2015) or that the AV is operated by the person in the right-hand seat (Habibovic et al., 2016; Habibovic et al., 2018). Therefore, the method has been recommended to examine the AV–pedestrian interaction (Rothenbücher et al., 2016).

(a) **(b)**

Figure 3.1 Wizard of Oz approaches: **(a)** seat cover (Fuest, Michalowski et al., 2019, p. 3955); **(b)** right-hand steered vehicle (Habibovic et al., 2016, p. 33).

The WoZ approach has the advantage that pedestrians behavior can be observed in a natural environment in contrast to images, videos or VR setups (Rothenbücher et al., 2016). This contrasts with the high cost of resources, such as the vehicle, the objective data measurement, a trained driver, and if necessary the test track (Fuest, Schmidt, & Bengler, 2020). Another disadvantage of WoZ studies in real traffic is the low reliability as confounders cannot be controlled, as it is not possible to recreate the same scenario for all participants (Deb et al., 2017). This can be prevented if studies are conducted on test tracks. However, it is difficult for the driver under the seat cover to stop at the same point in every trial (Moore et al., 2019) and it is unclear to what extent a driver can replicate driving profiles. Since it is a real vehicle with a driver, the method is not risk-free.

In recent years, the method has gained more popularity and is used to assess driving behavior (e.g., Currano et al., 2018; Fuest, Michalowski et al., 2018; Fuest, Michalowski et al., 2019; Moore et al., 2019; Rodríguez, 2017; Rothenbücher et al., 2016) and eHMIs (Clamann et al., 2017; Habibovic et al., 2018; Hensch et al., 2020b; Joisten et al., 2020; Mahadevan et al., 2018; Matthews et al., 2017).

3.6 Experimental Vehicle

Since AVs are still under development and driving on public roads without a (safety) driver is still prohibited by law (Federal Ministry for Digital and Economic Affairs, 2020), the evaluation of communication by using "real" AVs is difficult to realize. Besides WoZ vehicles, experimental vehicles can be used in real traffic and on test tracks to design communication. This approach requires an experimental vehicle and technical expertise to implement the driving profiles and eHMIs. Therefore, this method is very cost-intensive. In addition, also on test tracks, a safety driver is needed to intervene in an emergency situation in order to not endanger participants. Since it is not foreseeable whether pedestrians as participants would spontaneously step onto the road in order to test the AV, communication with pedestrians has so far only been tested rudimentarily at low speeds (5.4 km/h (Burns et al., 2019), or 20 km/h (Miguel et al., 2019)). Experimental vehicles were also used to evaluate the interaction between AVs and drivers (Potzy, Feinauer et al., 2019; Potzy, Feuerbach, & Bengler, 2019). Comparatively low speeds of 30 km/h and 50 km/h were chosen for these studies and both were conducted on a test track (Potzy, Feinauer et al., 2019). As a backup, a safety driver was in the vehicle (Potzy, Feinauer et al., 2019). Just as for the WoZ method, reliability is low as there is no controlled environment for real traffic studies.

Driving behavior (e.g., Potzy, Feinauer et al., 2019; Potzy, Feuerbach, & Bengler, 2019) and eHMIs (e.g., Burns et al., 2020; Miguel et al., 2019) can be assessed using experimental vehicles.

3.7 Summary of Methods

The advantages and disadvantages of the methods influence the choice of method depending on the research question. As this thesis focuses on the AV's driving behavior, images are not further examined. Images are unsuitable for the evaluation of driving profiles and rather used for the pre-selection of eHMIs (section 3.1). In addition, driving simulator studies are excluded for the thesis, as these studies focus on the AV’s communication with drivers and not with pedestrians. Furthermore, an experimental vehicle was not available. Thus, videos, VR and WoZ setups were used in this thesis.

In addition, the study results presented in chapter 2 give the impression that the choice of method affects the results. Especially in video studies, it is possible that driving behavior cannot be clearly perceived by pedestrians and eHMIs reduce this perception effect (Petzoldt et al., 2018). This assumption is supported by the fact that most positive effects for eHMIs were found in video and VR studies, whereas in WoZ studies eHMIs have a less positive influence on the pedestrians’ willingness to cross the road (Fuest, Schmidt, & Bengler, 2020). In contrast, it can be seen that the pedestrians crossing behavior is influenced by the driving behavior of the AV across all methods (Fuest, Schmidt, & Bengler, 2020).

In order to ensure the comparison of the study results, it is necessary to evaluate whether these differences are independent of the chosen method as it is otherwise difficult to derive conclusions about driving behavior or eHMIs. While first research findings suggest that VR and WoZ study results are valid in comparison to reality, there are no studies reviewing the validity of videos in the field of automated driving. However, as results differ, it remains unclear whether perception in VR and videos varies from the perception in reality. Therefore, a comparison between videos, VR, and WoZ setups were conducted to fill this research gap and to give recommendations which method should be used to evaluate the communication of an AV. Since different types of videos exist (section 3.2), two different types of videos were used for the comparison of methods: videos showing a WoZ vehicle and videos showing a VR vehicle.

Table 3.1 Advantages and Disadvantages of Methods.

Methods	Advantages	Disadvantages
Image	• cost-efficient (Rice et al., 2017; Wright, 2005) • usable for an online survey (Rice et al., 2017; Wright, 2005) • high reliability of laboratory studies • little knowledge required to edit images • risk-free • used for evaluating: o early-stage eHMI concepts (e.g., Fridman et al., 2019; Yang, 2017) o marked AVs (e.g., Hagenzieker et al., 2020) o different vehicle types (e.g., Dey et al., 2019) o driver state (e.g., Lundgren et al., 2017)	• perception may vary from the perception in reality • low validity (Farmer, 1998) • low reliability of online studies • evaluation of driving behavior only in a rudimentary manner (Fuest, Maier et al., 2020) • only for an initial impression (Fuest, Schmidt, & Bengler, 2020) • not immersive • evaluation with use case (with surrounding traffic, distraction or visual coverings) not possible
Video	• cost-efficient • usable for an online survey • high reliability of laboratory studies • driving behavior can be visualized • risk-free • used for evaluating: o eHMI concepts (e.g., Chang et al., 2018; Eisma et al., 2020; Li et al., 2018; Petzoldt et al., 2018; Song et al., 2018) o driving behavior (e.g., Ackermann et al., 2018; Dey et al., 2019; Fuest, Schmidt, & Bengler, 2020)	• perception may vary from the perception in reality (Petzoldt et al., 2018) • low reliability of online studies • less immersive
Virtual Reality	• relatively cost-efficient: technology must be purchased once (Burns et al., 2020; Deb et al., 2017) • immersive (Deb et al., 2017; Löcken et al., 2019) • high reliability (Deb et al., 2017) • .valid method to evaluate road crossing behavior (Deb et al., 2017) • easily transported and reinstalled elsewhere (Weber et al., 2019) • risk-free (Burns et al., 2020; Deb et al., 2017) • used for evaluating: o eHMI concepts (e.g., Böckle et al., 2017; Clercq et al., 2019; Holländer, Colley et al., 2019; Hudson et al., 2019)	• perception vary from the perception in reality (Bhagavathula et al., 2018; Schneider et al., 2021) • can lead to simulator sickness (Deb et al., 2017)

Methods	Advantages	Disadvantages
	o driving behavior (e.g., Burns et al., 2020; Dietrich, Maruhn et al., 2020; Fuest, Maier et al., 2020; Jayaraman et al., 2018; H. Schmidt et al., 2019)	
Driving Simulator	• highly immersive (Kantowitz, 2011) • risk-free • high reliability • used for evaluating: o eHMI concepts (e.g., Rettenmaier et al., 2019; Rettenmaier, Albers, & Bengler, 2020) o marked AVs (e.g., Fuest, Feierle et al., 2020; Kühn et al., 2019) o driving behavior (e.g., Kauffmann, Naujoks et al., 2018; Kauffmann, Winkler et al., 2018)	• cost-intensive(technology, license, technical support) • perception of speed vary from the perception in reality (Hurwitz et al., 2005) • familiarization phase necessary (Greenberg & Blommer, 2011) • can lead to simulator sickness (Allen et al., 2011)
Wizard of Oz	• can be used to simulate an AV (Dietrich, Boos et al., 2020; Habibovic et al., 2016; Habibovic et al., 2018; Rothenbücher et al., 2015) • evaluation in real traffic (Rothenbücher et al., 2016) and on test tracks • used for evaluating: o eHMI concepts (e.g., Clamann et al., 2017; Habibovic et al., 2018; Hensch et al., 2020b; Joisten et al., 2020; Mahadevan et al., 2018; Matthews et al., 2017) o driving behavior (e.g., Currano et al., 2018; Fuest, Michalowski et al., 2018; Fuest, Michalowski et al., 2019; Moore et al., 2019; Rodríguez Palmeiro, 2017; Rothenbücher et al., 2016)	• cost-intensive(vehicle, measurement of objective data, trained driver, and if necessary a test track) (Fuest, Schmidt, & Bengler, 2020) • low reliability (in a in real traffic study) (Deb et al., 2017) • reproducibility of driving behavior (Moore et al., 2019) • not risk-free
Experimental Vehicle	• evaluation in real traffic and on test tracks • used for evaluating: o eHMI concepts (e.g., Burns et al., 2020; Miguel et al., 2019) o driving behavior (e.g., Potzy, Feinauer et al., 2019; Potzy, Feuerbach, & Bengler, 2019)	• cost-intensive (e.g., vehicle, test track, safety driver) • (safety) driver as backup (Federal Ministry for Digital and Economic Affairs, 2020) • technical expertise necessary • low reliability (in a in real traffic study) • not risk-free • studies so far only at low speeds (Burns et al., 2019; Miguel et al., 2019; Potzy, Feinauer et al., 2019; Potzy, Feuerbach, & Bengler, 2019)

4 Metrics

Across all published studies evaluating the influence of communication between AVs and pedestrians, different metrics are used to compare driving profiles and eHMIs. Those metrics are based on objective and subjective data. Besides this, a distinction can be made between quantitative and qualitative data. Quantitative data are related to a measurable quantity and standardized (Przyborski & Wohlrab-Sahr, 2014). In contrast, qualitative data are related to the quality of the content and reconstructive (Przyborski & Wohlrab-Sahr, 2014). While quantitative methods empirically investigate hypotheses, qualitative methods are used to generate hypotheses or theories (Przyborski & Wohlrab-Sahr, 2014). As this thesis focuses on the investigation of hypotheses (chapter 6), this chapter emphasizes on objective and subjective quantitative data.

4.1 Objective Data

All metrics calculated from the objective data use a motoric reaction from pedestrians to compare AVs driving profiles or eHMIs. Therefore, the metrics can be assigned to the motor processor as part of the *Model Human Processor* (Figure 4.1; Card et al., 2008).

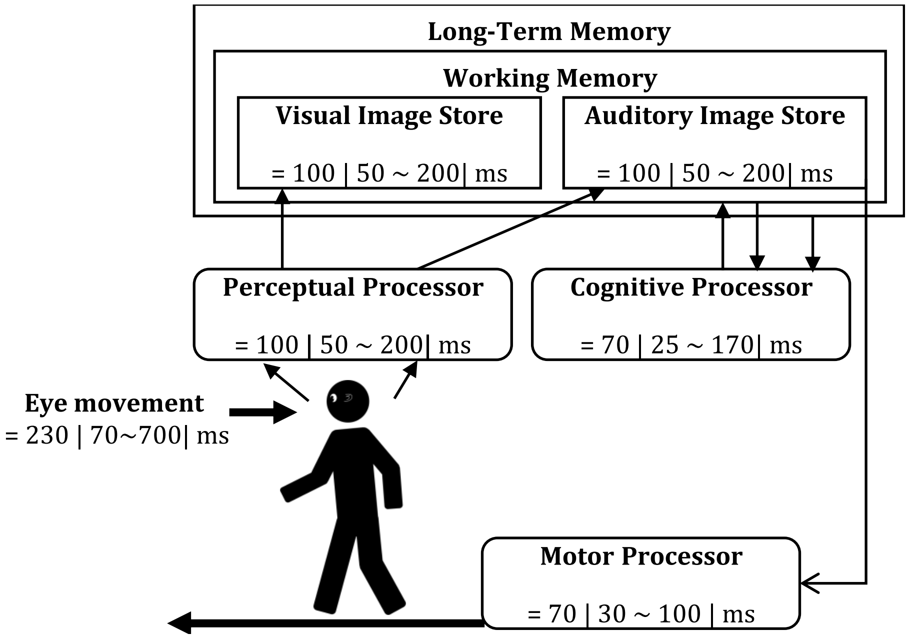

Figure 4.1 The Model Human Processor based on Card et al. (2008).

In the human being, a stimulus is recorded by the sensor systems of the body, such as the visual or auditory system. The stimulus is examined in the perceptual processor for known patterns. Therefore, beside the sensors, the perceptual system includes associated buffer memories, such as the visual and the auditory image store. Subsequently the stimulus is processed in terms of content in the working memory as well as in the cognitive processor and in the long-term. A motor process can subsequently take place. For all processes different amounts of time are needed, depending on the stimulus. (Card et al., 2008)

This process can be transferred to the pedestrian's decision to cross a road. The pedestrian perceives a stimulus from the traffic situation. However, the stimulus depends on the quality of the virtual and physical environments (Schneider & Bengler, 2020), such as daytime, e.g., dusk might impair the visual system (Ackermann, Beggiato, Bluhm et al., 2019; Beggiato et al., 2018). The pedestrian compares the stimulus with already experienced situations and based on the previous knowledge the best alternative action will be deduced. For example, the pedestrian perceives a decelerating vehicle and, based on previous experience, expects the vehicle to yield and therefore, to be able to cross the road safely. The output of those processes can be measured in different ways: using questionnaire items, pressing a button when a decision is made, or asking pedestrians to cross a road whenever feeling safe to do so.

In most studies, participants were asked to press and/or hold a button. For example, participants were asked to press the spacebar on a keyboard (Eisma et al., 2020) or a handheld button (Clercq et al., 2019) and hold the button as long as they felt safe to cross the road (Burns et al., 2019; Clercq et al., 2019; Eisma et al., 2020). The longer participants kept the button pressed, the better the eHMI (Eisma et al., 2020) or the driving profile. In another setup, participants were asked "to press a key at the last moment, when they would cross the street comfortably" (Beggiato et al., 2018, p. 1076). However, this cannot be used to analyze responses of participants to non-yielding AVs (Eisma et al., 2020). A non-yielding driving profile can also not be reviewed for studies in which the participants were asked to press the spacebar, as soon as they recognized a deceleration of the AV (Petzoldt et al., 2018). Asking the participants to press a button as soon as they recognize changes in the AV's driving profile (Ackermann et al., 2018; Ackermann, Beggiato, Bluhm et al., 2019) helps to analyze both driving profiles: yielding and non-yielding. After pressing the button, participants have to specify, whether they noticed the vehicle accelerated or decelerated (Ackermann et al., 2018; Ackermann, Beggiato, Bluhm et al., 2019). However, this metric cannot be used to evaluate a driving profile without speed changes.

Besides the question how long participants need to recognize changes in the AV's driving profile, other research focuses on the question whether and at what time pedestrians felt safe enough to cross the road. Therefore, different studies asked participants to press the spacebar (Song et al., 2018) or the button on the handheld controller (Chang et al., 2017) to indicate that they would cross the road, or even let the pedestrian cross the road (Clamann et al., 2017; Mahadevan et al., 2018; Miguel et al., 2019). The crossing time and error rate can be used to compare different eHMIs (Chang et al., 2017). Another possibility is the critical gap acceptance, where participants are asked to press a button (Beggiato et al., 2017; Joisten et al., 2020), raise their hand, or take a step forward or backward (Rodríguez Palmeiro, 2017) "at the last moment they are willing to cross the road in front of the vehicle" (Joisten et al., 2020, p. 253).

The crossing time (Deb, Strawderman, & Carruth, 2018; Dietrich, Maruhn et al., 2020; Jayaraman et al., 2018) and error rate (Holländer, Colley et al., 2019) were also used in VR studies in which participants crossed a virtual road by walking. Instead of the error rate, H. Schmidt et al. (2019) analyzed the success rate and the number of collisions. Deb, Strawderman, and Carruth (2018) additionally measured the waiting time, i.e., the time participants waited before entering the crosswalk, which is comparable to the crossing time used by Chang et al. (2017), and the decision time described by Holländer, Colley et al. (2019) and Holländer, Wintersberger, and Butz (2019), as well as the decision phase used by Clamann et al. (2017). The crossing time used by Deb, Strawderman, and Carruth (2018) refers to the time pedestrians needed to cross the whole road. This is comparable to the metrics used by Jayaraman et al. (2018) who used the waiting and crossing time, as well as a distance to collision, a jaywalking time and the crossing speed. This example illustrates that different names for the same metric exist. Therefore, a precise definition of the metrics used is essential for publications.

Dietrich, Maruhn et al. (2020) and Othersen et al. (2019) measured the *Crossing Initiation Time to Vehicle Stop* which refers to the time difference between the moment the AV stood completely and the pedestrian entering an encroachment zone. In the encroachment zone the AV's and the pedestrian's path would overlap (Dietrich, Maruhn et al., 2020). However, this metric can only be analyzed if the vehicle is going to yield. In addition, this metric, as well as the crossing time (Deb, Strawderman, & Carruth, 2018; Jayaraman et al., 2018) cannot be used in a WoZ setup to not endanger pedestrians when they mistakenly want to cross the road.

4.2 Subjective Data

Besides the objective data, subjective data were recorded during most studies. Participants answered questions regarding the perceived speed of the vehicle (Burns et al., 2020), the distance between the vehicle and the HRU (Burns et al., 2020), and their willingness to cross the road when a vehicle was approaching (Lundgren et al., 2017; Mahadevan et al., 2018). In addition, participants were asked to report the vehicle's intention (Chang et al., 2018; Moore et al., 2019; Zhang et al., 2018) or rate how certain they were about their decision to cross the road (Eisma et al., 2020). Participants rated how confident they felt with their interpretation of the intention (Habibovic et al., 2018) or how confident they felt while they crossed the road (Holländer, Colley et al., 2019). In another study, pedestrians rated their hesitations regarding their crossing behavior (Miguel et al., 2019). However, a road crossing cannot be implemented in every method (chapter 3). Therefore, some researchers cut videos and asked participants if they would cross the road at the moment the video stopped (Dey et al., 2017, 2019).

Questionnaire items with a Likert scale response format were used to evaluate eHMIs, such as asking for the recognizability, the unambiguousness, and the intuitive comprehensibility of the AV's message, to evaluate if participants understood the message correctly (Ackermann, Beggiato, Schubert, & Krems, 2019). Regarding implicit communication, participants assessed their confidence with the vehicle's behavior (Holländer, Wintersberger, & Butz, 2019) and the perceived effectiveness of the communication (Matthews et al., 2017). Moreover, subjective perceptions regarding perceived safety (Böckle et al., 2017; Burns et al., 2020; Chang et al., 2017; Deb, Strawderman, & Carruth, 2018; Habibovic et al., 2018; Holländer, Wintersberger, & Butz, 2019; Joisten et al., 2020; Matthews et al., 2017; Rodríguez Palmeiro, 2017; H. Schmidt et al., 2019), trust in the vehicle (Burns et al., 2020; Hensch et al., 2020b; Holländer, Wintersberger, & Butz, 2019; Jayaraman et al., 2018; Matthews et al., 2017; Nuñez Velasco et al., 2019; H. Schmidt et al., 2019), interaction comfort (Ackermann, Beggiato, Schubert, & Krems, 2019; Böckle et al., 2017), and interaction quality (Moore et al., 2019) were measured.

In addition, researchers asked participants if they recognized the vehicle as autonomous (if not introduced in the study setup) (Moore et al., 2019), about the positive and negative aspects of the eHMI (Burns et al., 2019), which eHMI they would prefer (Othersen et al., 2019), and what changes they would make (Othersen et al., 2019).

4.3 Summary of Metrics

Different objective and subjective data are used to evaluate different eHMIs and driving profiles. A motoric reaction is the basis for the objective data. Referring to the information process, it can be assumed that the different metrics need more or less time. The cognitive process, as described in the *Model Human Processor*, might be shorter for recognizing changes in the driving behavior (e.g., Ackermann et al., 2018; Ackermann, Beggiato, Bluhm et al., 2019) and longer for starting to cross the road (e.g., Chang et al., 2017; Holländer, Colley et al., 2019; Song et al., 2018), or even crossing the road completely (e.g., Deb, Strawderman, & Carruth, 2018; Jayaraman et al., 2018). A reason might be that stepping on the road while a vehicle approaches has a higher safety risk and pedestrians therefore need to be more confident in their decision. As a result, it is not possible to compare the absolute values of the metrics from different studies. Moreover, not all metrics can be used to evaluate a non-yielding intention or are suitable to be used in a WoZ setup.

In addition, a variety of questionnaire items are used to let participants rate, e.g., the driving behavior or the criticality of the experienced situations. However, a clear recommendation for appropriate metrics or questionnaire items cannot be derived from the study results.

5 Derivation of Metrics

The aim of the thesis is to give recommendations which method should be used to evaluate an AV’s driving behavior to communicate a yielding or non-yielding intention to a pedestrian. Therefore, a metric is needed that can be used for video, VR and WoZ setups. In addition, the metric is intended to be suitable for comparing both yielding and non-yielding intentions. It should also be possible to evaluate driving profiles with constant speeds. The following metrics cannot be used, due to these requirements. Pressing a button as long as pedestrians felt safe to cross the road (Burns et al., 2019; Clercq et al., 2019; Eisma et al., 2020) cannot be used as non-yielding intentions are not analyzable. Pressing a button when changes in the driving behavior are recognized (e.g., Ackermann et al., 2018; Ackermann, Beggiato, Bluhm et al., 2019) cannot be used due to the same reason. In addition, pedestrians crossing the road (e.g., Chang et al., 2017; Deb, Strawderman, & Carruth, 2018; Holländer, Colley et al., 2019; Jayaraman et al., 2018; Song et al., 2018) would endanger pedestrians in a WoZ setup and is therefore excluded.

The goal of designing AVs’ driving behavior should be to ensure that pedestrians can identify the intention of the AV as early as possible. This has the advantage that deadlock situations can be prevented. In addition, it can be assumed that pedestrians get used to communicating with AVs and adapt their crossing behavior over time. Pedestrians who recognize an intention early but do not cross the road because they feel unsure about AVs might adjust their behavior if they experience that AVs do not change their intention. Thus, at this early stage of research, it is useful not to examine current crossing behavior, but to evaluate when pedestrians recognize the intention of an AV. Therefore, a metric is needed that focuses on the time pedestrians need to recognize the AV's intention. The metric devised for this thesis is called *Intention Recognition Time* (IRT) and allows for the collection of objective data. In order not to endanger participants, they are asked to press a button at the moment at which they think they recognize the AV’s intention.

In addition, two other metrics are calculated using the moment participants think that they recognize the intention. First, the gap between the participant and the AV is calculated (chapter 9). Second, the driving data are used to analyze the time of decision. Therefore, it is calculated how many participants wait to press the button until the AV stands completely or passes by (chapter 11).

After pressing the button, participants are asked additional questionnaire items. The first item enquires about the assumed AV’s intention. Using this question the misinterpretation rate of intentions can be calculated. It can be assumed that an ambiguous driving profile leads to a high misinterpretation rate. This would make it possible to detect ambiguous driving profiles and distinguish between ambiguous and unambiguous driving profiles. The definition of an unambiguous driving profile is that the vehicle’s intention can be correctly communicated to a pedestrian.

To analyze the criticality of a driving profile pedestrians mention whether they would cross the road at the moment they recognized a yielding intention. It poses a safety risk, if pedestrians would cross the road when the AV intended to go first (Fuest, Schmidt, & Bengler, 2020). Moreover, it is desirable for traffic flow that pedestrians understand the intention correctly at an early stage so that they cross the road before the AV stands completely. Besides the mentioned crossing behavior participants rate the perceived criticality of the situation by using the criticality scale from Neukum et al. (2008).

In addition, the subjective decision-making reliability is surveyed by asking participants to rate their certainty about the AV's intention (from very uncertain to very certain on a five-point Likert scale). Besides having a safe and accident-free driving behavior, AVs should also be accepted by other HRUs. The AV's driving behavior should therefore be evaluated in detail. Participants are thus asked to evaluate the driving behavior (from very poor to very good on a five-point Likert scale).

The IRT and the questionnaire items were applied across all studies for the thesis and adapted if results suggest changes. Based on the study results, recommendations for the metrics and questionnaire items used were derived for further research.

6 Objectives and Contributions

The driving behavior is an essential feature of AVs as it is visible for HRUs as long as there is no visual impairment and, in accordance with the principle "you cannot not communicate" (Watzlawick, 2016), driving behavior is a permanent means of communication. In addition, the overview in chapter 2 suggests that the choice of method might influence the results of evaluating the AV's communication. Besides the variety of methods (chapter 3), research also relies on a number of different metrics (chapter 4). It remains unclear as to what extent the results of the existing research can be compared, due to the varying methods and metrics employed. Therefore, this thesis aims to determine which method is best suited to evaluate AV's driving behavior. To this end, video, VR and WoZ setups were compared using the devised IRT metric and questionnaire items (chapter 5).

In order to compare different setups, a traffic situation must be constructed whose attributes can be kept constant across all studies. Therefore, a taxonomy of traffic situations for the interaction between AVs and HRUs was created (chapter 7) and a suitable situation with different value facets was selected. This allowed for comparing the methods under exclusion of further disturbing factors. The situation was reviewed in an initial VR study and value facets were adjusted because of the results (chapter 8). Additionally, the metrics and the created questionnaire were assessed and an item was modified due to the findings (chapter 8). Subsequently, it was checked whether metrics and questionnaire can be implemented in a WoZ setup and whether the situation used in VR can be adapted to reality (chapter 9). In addition, it was necessary to examine to which extent driving profiles can be reproduced by a human driver (chapter 10). The results of these studies were considered in the comparison of different setups (chapter 11). Quality standards for the methods were collected across all studies in order to minimize possible weaknesses of the methods.

In summary, the following research questions are examined in this thesis:

- What methods can be recommended to evaluate an AV's driving behavior? Which quality standards must be fulfilled by the methods and what are possible weaknesses of the methods?
- Can the devised metrics be recommended to distinguish between different driving profiles to communicate an AV's yielding or non-yielding intention?
- What driving behavior can be recommended for AVs to communicate a yielding or a non-yielding intention to a pedestrian?

7 Taxonomy of Traffic Situations for the Interaction between Automated Vehicles and Human Road Users

Fuest, T., Sorokin, L., Bellem, H., & Bengler, K. (2018). Taxonomy of Traffic Situations for the Interaction between Automated Vehicles and Human Road Users. In N. A. Stanton (Ed.), *Advances in Intelligent Systems and Computing. Advances in Human Aspects of Transportation* (Vol. 597, pp. 708–719). Cham: Springer International Publishing.

As this thesis aims to compare different methods that are used to evaluate AV's driving behavior, a traffic situation is needed whose attributes can be kept constant across all studies. Therefore, a taxonomy was created that structures traffic situations in attributes and related value facets that influence the interaction between AVs and HRUs. Since there is a large variety of situations in road traffic (Klebelsberg, 1982), the taxonomy also helps to answer the question whether the designed and evaluated communication can be transferred to other traffic situations.

The taxonomy is clustered in 12 attributes (see left column of Table 7.1.). All attributes have belonging value facets. Two of the attributes of the taxonomy are the right of way and the AV's intention regarding the right of way. Both attributes are chosen since they influence the expression of the AV (ask for the right of way vs. enforce the right of way) and might influence the HRU's acceptance of the AV. The belonging value facets are: AV has the right of way, HRU has the right of way or right of way is undefined. The latter can be found in parking or shared space areas. Other attributes are, e.g., the HRU character (e.g., pedestrians, bicyclists, and drivers) and their speed (0 km/h, 4.4 km/h, 17.5 km/h, 30 km/h, 50 km/h, and 130 km/h).

Based on the taxonomy, a situation for this thesis was built by defining all relevant value facets for the attributes (see right column of Table 7.1). A pedestrian was chosen as HRU character because the thesis focuses on the communication between AVs and pedestrians. In addition, the right of way and the AV's intention were varied to evaluate if the AV's driving behavior should differ in various right of way situations (chapter 8). In addition, different speeds were used (30 km/h and 49 km/h; chapter 8 and chapter 9). Besides these attributes, all other attributes were kept constant across all studies to reduce potential confounders. During all studies, the view of the AV was unobstructed at all times and the pedestrian stood at the roadside while the AV approached from left.

Table 7.1 Overview of chosen attributes and value facets for the studies of the thesis.

Attribute	Value Facets
Right of way	AV / HRU / Undefined
AV's intention regarding right of way	Let the HRU go first / AV goes first
HRU's intention regarding right of way	Not defined
HRU Character	Pedestrians
Longitudinal distance (Headway)	> 10 Meters
Lateral distance	≤ 3 Meters
Attention HRU	Yes
Impairment of the HRU's perception	No impairment
Speed AV	30 km/h / 49 km/h
Speed HRU	0 km/h
Driving direction AV	Driving forward
Perspective (from the perspective of the AV)	Ahead

8 How Should an Automated Vehicle Communicate Its Intention to a Pedestrian? – A Virtual Reality Study

Fuest, T., Maier, A.S., Bellem, H., Bengler, K. (2020). How Should an Automated Vehicle Communicate Its Intention to a Pedestrian? – A Virtual Reality Study. In: Ahram T., Karwowski W., Pickl S., Taiar R. (eds) *Human Systems Engineering and Design II. IHSED 2019. Advances in Intelligent Systems and Computing*, Vol 1026. Springer, Cham.

To further define the value facets, an initial VR study was conducted to determine whether the right of way had an effect on the metrics and questionnaire items. Using the software Unity, three situations were implemented with different right of way situations in an urban environment. In order to let the AV have the right of way a single-carriageway with two lanes but without road marking was used in the first scenario. To give the HRU the right of way a zebra crossing was added to the road in the second scenario. In the third scenario, a parking area was used in a shared space situation, so that the right of way was undefined. The AV maintained a constant speed of 49 km/h to communicate the non-yielding intention. For the yielding intention the AV decelerated in two different ways. In the first driving profile, the AV started to decelerate early at a 25 m distance from the pedestrian's position with at maximum -6 m/s^2. In the second driving profile, the AV started to decelerate late at a 16 m distance from the pedestrian's position with at maximum -11 m/s^2. The driving profiles were implemented by changing the angle of the brake pedal in percentages.

The results showed a significantly better rating of the driving behavior if the AV had the right of way. This was due to the fact that pedestrians perceived a speed of 49 km/h in a parking area as too fast. No significant differences between the different rights of way were found for the other dependent variables. Therefore, the AV's speed was adapted to 30 km/h and the right of way was kept undefined for further studies. In addition, recommendations for AV's driving profiles were derived. If the AV yields, pedestrians preferred that the deceleration starts at a 25 m distance from the HRU's position, instead of a 16 m distance. The AV should come to a standstill at 5 m from the HRU's position. If the AV insists on its right of way it should maintain a constant speed.

The study further aimed at testing the setup and metrics. The study showed that the criticality scale of Neukum et al. (2008) is not applicable for assessing the perceived criticality of the situation from a pedestrian's perspective. Participants mentioned problems with the anchor points (e.g. "not controllable") and in addition, the results showed a large deviation in answers. Therefore, the variable was adjusted and a five-point Likert scale (from very critical to very uncritical) as response format was used for further studies. In addition, as already mentioned in section 3.3, the perception in VR deviates from real values for the estimation of speed (Bhagavathula et al., 2018). This can be confirmed by the study results: a deceleration of -6 m/s^2 and even -11 m/s^2 still looked like a "normal" braking maneuver. However, comfortable braking for vehicle passengers is at a maximum deceleration rate of -3.4 m/s^2 (AASHTO, 2011). Therefore, the results suggest more realistic driving profiles should be implemented for a VR setup. A prerequisite for a VR simulator should therefore be that driving data can be implemented properly and not by changing the angle of the brake pedal in percentages.

9 Using the Driving Behavior of an Automated Vehicle to Communicate Intentions – A Wizard of Oz Study

Fuest, T., Michalowski, L., Träris, L., Bellem, H., & Bengler, K. (2018). Using the Driving Behavior of an Automated Vehicle to Communicate Intentions – A Wizard of Oz Study. In *21st International Conference on Intelligent Transportation Systems* (ITSC) (pp. 3596–3601). IEEE.

Building on the previous results (chapter 8), a shared space situation was used to evaluate whether the AV's driving behavior can be used to communicate intentions to a pedestrian in a WoZ setup. Longitudinal and lateral driving parameters were defined to communicate one yielding and two non-yielding intentions. Based on human driving behavior, the AV decelerated with at maximum -2.4 m/s^2 starting at a 22 m distance from the pedestrian position to communicate the yielding intention. For both non-yielding driving profiles, the AV maintained its speed. The difference between the two driving profiles was an increased lateral distance of 50 cm towards the center of the road. Since one finding from the previous study was that pedestrians expect a reduced speed in a parking area (chapter 8), the AV drove with a maximum speed of 28.5 km/h (30 km/h indicated on the speedometer). To simulate an AV, the driver was hidden from the pedestrians' view by sitting under a seat cover (Rothenbücher et al., 2016). However, the driver used the seat cover in only half of all trials; therefore, it was possible to investigate whether the visibility of the driver changes the IRT and questionnaire items. In order to assess the driving behavior, the participants were asked to rate the perceived criticality on a five-point Likert scale (chapter 8).

Results showed that the visibility of the driver had no influence on the dependent variables. Therefore, it can be assumed that, at a vehicle speed of 28.5 km/h, pedestrians focus more on the driving behavior rather than on a possible communication with the driver. The misinterpretation rate was low for all driving profiles. Therefore, pedestrians confirmed the understandability of all driving profiles used for the study. As the previous study showed, the AV should maintain a constant speed if it insists on its right of way. However, an additional lateral movement of 50 cm towards the center of the road increased the participants' evaluation of the driving behavior and reduced their perceived criticality of the situation.

The study further showed that the IRT can be applied not only in a VR setup (chapter 8) but also in a WoZ setup. However, the IRT and the misinterpretation rate showed no significant differences between the different driving profiles used for the study. It can be assumed that the IRT is not sensitive enough to differentiate between two unambiguous driving profiles. However, the pedestrians' evaluation of the driving behavior and the rating of the perceived criticality were useful for selecting the preferred and thus least critical driving behavior.

The following open issues can be derived from the results. Firstly, it should be verified whether participants detect the seat cover. Secondly, it can be assumed that a driver is not able to drive a target driving profile without deviations. Therefore, it is also important to examine the capability of a human driver to replicate driving profiles.

10 Reproducibility of Driving Profiles – Application of the Wizard of Oz Method for Vehicle Pedestrian Interaction

Fuest, T., Michalowski, L., Schmidt, E., & Bengler, K. (2019). Reproducibility of Driving Profiles–Application of the Wizard of Oz Method for Vehicle Pedestrian Interaction. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)* (pp. 3954-3959). IEEE.

In order to investigate whether participants detect the seat cover and how accurate a non-professional driver can follow predefined driving profiles, a WoZ study was conducted. Based on the previous findings, unambiguous and ambiguous driving profiles for the intentions “Let the HRU go first” and “AV goes first” were used to evaluate whether IRT is sensitive enough to differentiate between these strongly disparate driving profiles. The AV’s maximum speed was 28.5 km/h (30 km/h indicated on the speedometer). This speed was maintained for the “AV goes first, unambiguous” driving profile. For the driving profile “AV goes first, ambiguous” a changing speed profile was used. The speed was reduced to 15 km/h (indicated on the speedometer) at a 40 m distance from the pedestrian’s position. The driver accelerated again to the initial speed at a 7.4 m distance. An early acceleration was used for the “Let the HRU go first, unambiguous” driving profile: the driver decelerated with at most -1.5 m/s^2 starting at a 40 m distance from the pedestrian’s position. The vehicle stood completely at a 7.4 m distance from the pedestrian’s position. For the driving profile “Let the HRU go first, ambiguous” the driver decelerate with at most-4.1 m/s^2 starting at a 14.8 m distance from the pedestrian’s position. A complete stop was reached at a 4.3 m distance from the pedestrian’s position.

The wizard was a non-professional driver who managed to stay within a tolerance limit of ±2 km/h in 84% of the trials. While it was easy for the driver to maintain speed using a speed limiter, it was difficult to slow down smoothly over a long period. The most difficult part for the driver was to strongly decelerate up to a defined stop line. Therefore, more complex driving profiles are harder to follow.

Results showed that over 85% of the participants believed that they encountered a real AV. The seat cover is therefore useful to simulate an AV from the pedestrian perspective. Comparing ambiguous and unambiguous driving profiles, the IRT revealed only significant differences for the intention “AV goes first”, but not for the intention “Let the HRU go first”. It can be assumed that these differences occur due to the design of the driving profiles. If the AV brakes late, the intention can only be recognized late, as compared to an early braking. Therefore, participants needed more time to correctly recognize the intention.

The results of the study suggest that the AV should change the driving profile in a range from 25 m to 8 m distance from the pedestrians’ position to communicate intentions as most pedestrians make their decision in this range. The finding of the previous studies that maintaining speed communicates best that the AV is not going to yield is confirmed by the results. Pedestrians prefer if the AV brakes early when yielding. The AV should start to decelerate 40 m away from the pedestrian with at maximum -1.5 m/s^2 to come to a complete stop 7.4 m before the pedestrian’s position.

11 Comparison of Methods to Evaluate the Influence of Automated Vehicle's Driving Behavior on Pedestrians: Wizard of Oz, Virtual Reality, and Video

Fuest, T., Schmidt. E., & Bengler, K. (2020). Comparison of Methods to Evaluate the Influence of Automated Vehicle's Driving Behavior on Pedestrians: Wizard of Oz, Virtual Reality, and Video. *Information* 11 (6), 291.

In order to derive which method can be used to evaluate the AV's driving behavior four study setups were compared: a WoZ, a VR setup, and two video setups (one using videos from the VR setup and one using videos from the WoZ setup). The driving profiles used are described in chapter 10. In real traffic, pedestrians not only have to recognize the vehicle's intention, but also to make a decision whether to cross the road. Therefore, the IRT was compared with the metric of the crossing behavior. This comparison was only made for the VR setup because pedestrians cannot walk in a video setup and could be endangered in the WoZ setup.

Results demonstrate that the misinterpretation rate was higher for ambiguous driving profiles, as compared to unambiguous driving profiles for all four setups. Therefore, the metric can be used to distinguish between unambiguous and ambiguous driving profiles for all setups. However, a detailed analysis of the misinterpretation rate revealed that it is lower for the VR and the video setups, as compared to the WoZ setup. A similar result was shown for the pedestrians' mentioned hypothetical crossing behavior. In the WoZ setup, fewer pedestrians would have crossed the road at the moment they recognized the AV's yielding intention, as compared to the other setups. Differences between the video setups were also evident, as by far the most participants would have crossed the road in the video VR setup for the unambiguous driving profile. The results for the mentioned hypothetical crossing behavior for the intention "AV goes first, ambiguous" also depended on the type of video: the risk of crossing is underestimated in the video VR setup just as in the VR setup, whereas the video WoZ setup could reproduce the results of the WoZ setup. In addition, the IRT led to different results across all four setups. The metric was able to distinguish between unambiguous and ambiguous driving profiles in both video setups. In contrast, in the WoZ and VR setup, differences were only evident for the intention "AV goes first" but not for the intention "Let the HRU go first". To sum up, the results demonstrate that there are differences between the study setups. Particularly for the video setups perception and decision artefacts might occur.

Results for the comparison of the IRT with the crossing behavior suggest that pedestrians assessed driving behavior early, but waited to cross the road until they felt confident about their decision. The longer waiting times resulted in pedestrians observing the whole driving profile. This led to more correct decisions regarding the AV's intention, as compared to the IRT metric. However, since two participants mistakenly crossed the road anyway, the metric is not appropriate for the WoZ setup.

12 General Discussion

In the following chapter, the study results are discussed in order to derive recommendations for the methods (section 12.1), metrics (section 12.2), and the AV's driving behavior (section 12.3). In addition, the taxonomy is used to discuss as to which extent the results for driving behavior can be transferred to other situations (subsection 12.3.3).

12.1 Comparison of the Methods

A total of six studies—two video, two VR and two WoZ studies—were conducted for this thesis. However, in order to compare the methods the same study design was used in four of the six studies. These four setups consisted of a WoZ setup, a VR setup, and two video setups, one with videos from the VR setup and one with videos from the WoZ setup (chapter 11). Results of the dependent variables (chapter 5) were compared to analyze which setup is best to evaluate an AV's driving behavior. In the following, the results of the studies are discussed in relation to findings from other studies.

To fulfill the quality criterion of objectivity (chapter 3), the instruction and the questionnaire were standardized across the studies. In addition, the wizard practiced the driving profiles several times before each study and roadside markings helped with orientation when to brake or accelerate. Moreover, the same wizard was employed across all trials and studies.

12.1.1 Wizard of Oz

An AV can be simulated by using a seat cover. Results revealed that pedestrians who were not familiar with the technology of AVs did not recognize that a driver was hidden under the seat cover (Fuest, Michalowski et al., 2019). This underlines the results from other studies that recommend the use of a WoZ vehicle (Dietrich, Boos et al., 2020; Rothenbücher et al., 2015, 2016). Therefore, the WoZ setup is valid to simulate an AV.

A WoZ approach is recommended for detecting ambiguous from unambiguous driving profiles (Fuest, Schmidt, & Bengler, 2020). The WoZ setup is suitable to evaluate driving behavior as long as the driving profiles are not too similar and differ in at least 6 km/h over the entire driving profile. This results from the fact that a non-professional driver requires a tolerance range of ±3 km/h for each driving profile (Fuest, Michalowski et al., 2019). For the comparison of more similar driving profiles the use of an experimental vehicle is recommended.

The reliability depends on the target driving profile, since the wizard can reproduce different driving profiles to different degrees of success. A speed limiter should be part of the equipment of the WoZ vehicle as it supports the driver to accelerate in a similar way over all trials and to maintain speed during the trial (Fuest, Michalowski et al., 2019). This is confirmed by the result that that a driver has problems to maintain speed across different trials and conditions if no driver assistance system is implemented in the vehicle (Rodríguez Palmeiro, 2017). In contrast, the WoZ approach cannot be fully recommended when a smooth deceleration over a longer distance and a deceleration to a standstill at a given distance are examined (Fuest, Michalowski et al., 2019). The latter is in line with the results by Moore et al. (2019) who observed that the driver had difficulties braking at a specific point in all trials. Therefore, it is important to record the driving data during the study for control purposes.

Nonetheless, the WoZ setup is most similar to a real scenario. Therefore, it is assumed that the results for the WoZ setup are the most realistic values. In the following, the other setups are compared to the values of the WoZ setup due to this fact.

12.1.2 Virtual Reality

VR can be used to detect ambiguous driving profiles. This is based on the fact that the misinterpretation rate is higher and the crossing rate is lower for ambiguous driving profiles, as compared to unambiguous driving profiles (Fuest, Schmidt, & Bengler, 2020). However, the misinterpretation rate and the collision risk for the ambiguous driving profiles are underestimated in VR, as compared to the WoZ setup (Fuest, Schmidt, & Bengler, 2020). This means that ambiguous driving profiles are perceived as less ambiguous than they are in reality. In contrast, the mentioned hypothetical crossing behavior when the AV intended to yield is overestimated in VR, as compared to the WoZ setup. This contrasts with the results of Bhagavathula et al. (2018) who found that pedestrians’ crossing intentions did not differ between VR and real environment situations. However, the setup was different: participants were asked to assess whether they would cross the road at a defined distance of 47.2 m between them and the vehicle (Bhagavathula et al., 2018). The vehicle maintained a constant speed of 40.2 km/h and did not decelerate (Bhagavathula et al., 2018). This suggests that participants are able to estimate distance in VR as well as in reality. That presumption is confirmed by the results of Iryo-Asano et al. (2018) who observed that the perception of distance in VR did not differ from the perception in real world situations. However, if not only the distance but also the deceleration needs to be noticed and interpreted by the pedestrians the results between VR and reality differ: intentions were recognized later in VR due to perceptual differences, as compared to reality (Fuest, Schmidt, & Bengler, 2020). This leads to the assumption that the deceleration is difficult to estimate in VR. The differences in perception are supported by the results of the second study (chapter 8). Strong deceleration values of $-6\ m/s^2$ and even $-11\ m/s^2$ were perceived as a normal braking maneuver through the head mounted display (Fuest, Maier et al., 2020). Schneider et al. (2021) emphasize this assumption by pointing out that the perceptual accuracy in VR does not correspond to perceptual accuracy in reality. In addition, it was shown that pedestrians understood the AV’s intention in the WoZ setup earlier than in the VR setup (Fuest, 2019). This is in line with the results of Maruhn et al. (2020) who stated that participants started to cross the road earlier in a real environment than in an augmented reality condition. Therefore, absolute values cannot be transferred to reality.

As the results indicate that perception of deceleration in VR differs from reality, it is important to publish the driving data used. For this purpose, it is essential that an implementation with specified driving parameters in the VR simulation is supported.

12.1.3 Videos

Just as for the WoZ and the VR setup, it is also possible for the video setups to distinguish between unambiguous and ambiguous driving profiles with the help of the misinterpretation rate and the crossing rate (Fuest, Schmidt, & Bengler, 2020). Nevertheless, the results are only transferable to a limited extent. The misinterpretation rate was lower in both video setups, as compared to the WoZ setup (Fuest, Schmidt, & Bengler, 2020). This results from the fact that more participants made their decision at an later stage, as compared to the WoZ setup and also as compared to the VR setup (Fuest, Schmidt, & Bengler, 2020). This suggests that perception and subsequent interpretation of driving profiles in a video setup is more difficult than in a VR setup, and also more difficult than in reality.

Moreover, the results show that the choice of video matters. as, for example, the time of decision and the collision risk are influenced by the video type (Fuest, Schmidt, & Bengler, 2020). However, a clear recommendation for one video type cannot be derived. Using videos from a WoZ setup would be recommended for investigating the crossing rate, as the crossing rates for both intentions when using videos from the WoZ setup are similar to the results from the WoZ setup (even if large deviations can be seen) (Fuest, Schmidt, & Bengler, 2020). This contrasts with the analysis of the time of decision for which videos from the VR setup are closer to the results from the WoZ setup, although not ideal (Fuest, Schmidt, & Bengler, 2020).

In addition, depending on the type of video, either a VR environment has to be implemented or a wizard has to drive a target driving profile, so the effort required for a video study is only slightly less than for other setups. For this reason and due to the perception and decision artefacts that occur in video setups, a video setup is not recommended for evaluating AV's driving behavior.

12.1.4 Limitations of the Comparison

The investigated methods were compared by replicating the same study design in four different setups. However, there are some aspects that could not be kept constant across the different setups.

In the VR setup, the situation was frozen and in the video setups, the videos stopped at the moment the participants pressed the button. This had the benefit that pedestrians were not influenced by the rest of the driving profile and could relate their answers only to what they had experienced. Implementing a freezing situation is much more difficult in the WoZ setup. Participants were asked to turn toward the experimenter as soon as they pressed the button. The experimenter stood facing away from the traffic situation, so that the participants also had to turn around. In addition, the wizard was notified of the intention detection by a visual signal and accelerated the vehicle to the initial speed, regardless of the original intention of the vehicle (Fuest, Schmidt, & Bengler, 2020). Participants were informed that reacceleration was independent of the experienced driving profile and should not be assessed. However, based on the *Model Human Processor* it can be assumed that a few milliseconds passed between the perception of the visual stimulus by the wizard and the renewed acceleration and that the participants therefore at least noticed the engine noise (Card et al., 2008; Rettenmaier & Bengler, 2020). Since reacceleration occurred only after participants pressed the button, IRT is independent of the influence of further driving. Nevertheless, it could have affected the subjective variables, such as the misinterpretation rate or the crossing rate. However, since the misinterpretation rates for the unambiguous driving profiles are comparable to those from the other setups and the misinterpretation rates for the ambiguous driving profiles in the WoZ setup are the highest (over 70%) across all setups, a possible influence seems to be very small.

Another limitation is the missing sound in the VR and video setups. It seems that pedestrians perceive a deceleration acoustically by the sound of the engine and the roll noises (Dietrich, Boos et al., 2020). However, in the studies published within this thesis, only one participant (<1% of all participants) mentioned that the deceleration was recognized by the AV shifting down to a lower gear. In addition, since most studies in VR and video setups are without sound, it can be concluded that pedestrians do not need the engine sound or roll noises to detect the AV's intention (e.g., Ackermann, Beggiato, Bluhm et al., 2019; Fuest, Maier et al., 2020; Fuest, Schmidt, & Bengler, 2020). Nonetheless, due to the lack of sound it might be that participants need more time to recognize the AV’s intention. This could explain why participants waited in the VR and video setups more often until the AV came to a standstill or passed by, as compared to the WoZ setup. However, this would not explain the differences between the VR and the video setups. It is more likely that perception in VR and videos is impaired and therefore participants need more time to recognize intentions in those setups, as compared to reality. This assumption is underlined by the results of Schneider et al. (2021) who indicate that the perceptual accuracy of VR simulators is insufficient to replicate real-world behavior.

It was not possible to compare the absolute values of the IRT between all study setups. This is due to the fact that the videos were recorded manually and the start of the video sequences was thus not synchronized. It is also difficult to compare the IRT for WoZ and VR setup because the driving profiles could not be fully replicated by the driver in the WoZ setup during all trials. However, the limitations for comparing the IRT in the WoZ and VR setup are minimal and the results are consistent with findings from other studies (Maruhn et al., 2020). Results showed that participants needed more time in the VR setup to understand the AV’s intention, as compared to the WoZ setup (subsection 12.1.2; Fuest, 2019). This is in line with the result that more participants waited to see the whole driving profile in VR, as compared to the WoZ setup. Transferring these results, it can be assumed that in the video setups participants needed even longer to recognize the AV’s intention. This is based on the fact that even more pedestrians waited in the video setups to see the whole driving profile. This suggests that the perception of driving profiles in videos is even worse than in a VR setup.

Further limitations of the method comparison are that type and color of the vehicle differed between the different setups. In the WoZ setup, a BMW 2 series (F46, 220d xDrive) in dark grey was used, in the VR setup, a BMW 3 series (F30) in silver was implemented and also presented in the videos, and for the videos from WoZ setup, a BMW 3 series (F31, BMW 320d Touring) in dark blue was used (Figure 12.1). The vehicle types were kept as constant as possible in the two video setups. The same color was not available for the simulation. For the WoZ setup, the vehicle had to be changed in order to record the driving data during the study. Based on previous studies only differences between large vehicles, such as a Ford F150 and a Mercedes Benz Van, and small and medium vehicles, such as a Smart, BMW Z4 or BMW i3, were found for the mean feel-safe percentages (Clercq et al., 2019) and for reaction time (Ackermann, Beggiato, Bluhm et al., 2019) whereas no differences between small and medium vehicles were found for both metrics (Ackermann, Beggiato, Bluhm et al., 2019; Clercq et al., 2019). In addition, Ackermann, Beggiato, Bluhm et al. (2019) used a Smart Electric Drive colored in white as a small vehicle and a BMW i3 in black as a medium-sized vehicle. As the results showed no differences for the reaction time between the different vehicles, it can be assumed that small differences in size and color of the vehicles used for the studies in this thesis (Figure 12.1) did not influence the results.


(a) (b) (c)

Figure 12.1 Vehicles used as AVs in the different setups: **(a)** BMW 2 series for the WoZ setup (Fuest, Michalowski et al., 2019, p. 3954); **(b)** BMW 3 series for the VR and the video VR setup (Fuest, Schmidt, & Bengler, 2020, p. 5); **(c)** BMW 3 series for the video WoZ setup (Fuest, Schmidt, & Bengler, 2020, p. 5).

Moreover, a parking area in an urban environment was used for the VR setup, whereas the parking area was in a rural environment in the WoZ setup (Figure 12.2). However, studies have shown that the level of abstraction of the environment does not influence the study results (Angerstorfer et al., 2019; Dönch, 2018). It can be inferred from the studies that a very abstracted environment or no environment at all produces no differences in results. Based on this, it can be assumed that a rural and an urban environment will also not lead to differences in the results. Nevertheless, further investigation might be useful.


(a) (b)

Figure 12.2 Pictures from the video setups: **(a)** Wizard of Oz; **(b)** virtual reality (Fuest, Schmidt, & Bengler, 2020, p. 5).

12.2 Recommendations for Metrics to Evaluate the Driving Behavior

Objective and subjective data were collected in order to analyze the results and to compare the methods. As described in chapter 6, the IRT is used as an objective metric. For the subjective data, five questionnaire items regarding the assumed intention, the mentioned crossing behavior, the subjective decision making reliability, the evaluation of driving behavior, and the perceived criticality were used.

12.2.1 Objective Metrics

The IRT was developed to distinguish between different driving profiles. The results of the first VR study showed significant differences between the driving profiles for the IRT (chapter 8, Fuest, Maier et al., 2020), whereas the results of the first WoZ study showed no differences (chapter 9, Fuest, Michalowski et al., 2018). While the driving profiles investigated in the VR study differed strongly, they were kept similar in the WoZ study (Table 12.1 and Table 12.2). This raised the suspicion that the metric is not sensitive enough to distinguish between unambiguous driving profiles, but can identify unambiguous from ambiguous profiles. However, this assumption could not be fully confirmed during a further verification (chapter 11, Fuest, Schmidt, & Bengler, 2020). The results varied between the different setups and AV's intentions (Fuest, Schmidt, & Bengler, 2020).

In addition, the percentages of participants who waited until the AV came to a complete standstill or passed by before pressing the button were analyzed. However, the results showed that the metric is not suitable to distinguish between unambiguous and ambiguous driving profiles. The results of the WoZ study showed that pedestrians—who correctly recognize the intention—made their decision before the vehicle stopped completely or passed by. This was independent of the unambiguity of the driving profile. (Fuest, Schmidt, & Bengler, 2020)

A reason for the fact that both metrics are not suitable to analyze AV's driving behavior is the design of the driving profiles themselves. The AV needs different amounts of time to drive the same distance due to the altered braking characteristics and therefore changing speeds (Figure 12.3). Since the IRT is a reaction time, results are not independent of the AV's driving behavior. Therefore, IRT is only suitable to a limited extent for comparing different driving profiles.

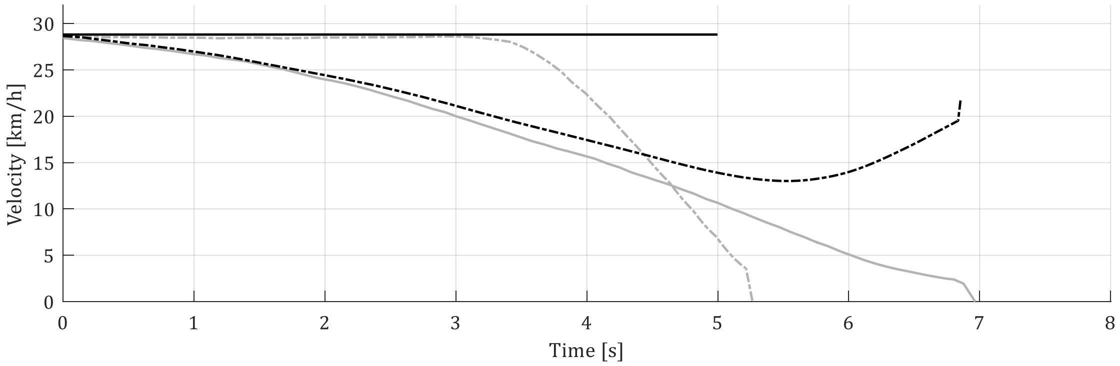

Figure 12.3 Unambiguous (solid lines) and ambiguous (dashed lines) target driving profiles for the intentions "Let the HRU go first" (gray lines) and "AV goes first" (black lines). The driving profiles end, when the AV came to a complete stop or passed the pedestrian.

This is also illustrated by the evaluation of the gap between AV and pedestrian at the moment the pedestrian understood the intention (Fuest, Michalowski et al., 2018). The problem applies to all metrics that require a time response by the participant to different driving profiles, e.g., the detection time (Ackermann et al., 2018; Ackermann, Beggiato, Bluhm et al., 2019), identification time (Petzoldt et al., 2018), and also to the crossing behavior. Consequently, no significant differences between the unambiguous and ambiguous driving profiles were found for the crossing behavior in the VR study (Fuest, Schmidt, & Bengler, 2020). The results showed that participants needed less time to correctly identify the AV's intention, as compared to the decision to actually cross the road (Fuest, Schmidt, & Bengler, 2020). This might due to the fact that the cognitive processor of the *Model Human Processor* (chapter 4) takes more time when crossing the street. A reason might be the higher safety risk when actually crossing, instead of pressing a button when recognizing the intention. In addition, also the motor processor might need more time when crossing the road. As the aim of a well-designed driving behavior is an early identification of the AV's intention to avoid deadlock situations and to give HRUs the opportunity to get used to the driving behavior (chapter 5), the IRT is a more appropriate metric to compare different driving profiles, as compared to other metrics. This is also due to the reason that the crossing behavior cannot be used to analyze a non-yielding intention of the AV.

12.2.2 Subjective Metrics

As the IRT alone does not seem to be useful for analyzing different driving profiles, subjective data are needed as a complement. Calculating the misinterpretation rate for each driving profile helps to distinguish between unambiguous and ambiguous driving profiles (Fuest, Maier et al., 2020; Fuest, Michalowski et al., 2018; Fuest, Michalowski et al., 2019; Fuest, Schmidt, & Bengler, 2020). Ambiguous and thus less understandable driving profiles lead to higher misinterpretations rates (Fuest, Maier et al., 2020; Fuest, Michalowski et al., 2018; Fuest, Michalowski et al., 2019; Fuest, Schmidt, & Bengler, 2020). However, the misinterpretation rate cannot be used to differentiate between two unambiguous driving profiles (Fuest, Michalowski et al., 2018).

For the VR setup, the misinterpretation rate was also calculated for the moment pedestrians started to cross the road. Here, no differences were found for the misinterpretation rate when comparing unambiguous and ambiguous driving profiles. By waiting longer to cross the road, pedestrians saw more of the full driving profile and therefore recognized the intention more often correctly. Nevertheless, two participants would have crossed the road by mistake, although the vehicle did not intend to yield. As a result, no participant should be asked in a WoZ or experimental vehicle setup to actually cross the road. (Fuest, Schmidt, & Bengler, 2020)

To analyze the safety-criticality of a driving profile, it was calculated how many pedestrians mentioned that they would cross the road at the moment they recognized the intention. Results showed a higher crossing rate for the unambiguous driving profile, as compared to the ambiguous driving profile, when the AV intended to yield (Fuest, Schmidt, & Bengler, 2020). In contrast, more pedestrians would have crossed the road by mistake when the AV did not intend to yield and drove ambiguously (Fuest, Schmidt, & Bengler, 2020). With the help of the mentioned crossing behavior, it is therefore possible to identify potentials and weaknesses of driving profiles.

To ensure that the driving profiles are perceived as non-critical, pedestrians rated the perceived criticality for the different driving profiles. In a first step, the situation criticality scale from Neukum et al. (2008) was used. Participants were asked to evaluate the situation on a scale from zero to ten described as “not noticeable”, “harmless”, “uncomfortable”, “dangerous”, and “not controllable” (Neukum et al., 2008). The scale has been developed to evaluate the criticality of situations from the driver’s perspective (Neukum et al., 2008). However, the pedestrians in the conducted studies stood at the roadside and had to recognize the intention of the vehicle (Fuest, Maier et al., 2020; Fuest, Michalowski et al., 2018; Fuest, Schmidt, & Bengler, 2020) or were asked to decide whether they would cross the road (Fuest, Schmidt, & Bengler, 2020). Therefore, they never felt that the situation was “not controllable” as they never lost the controllability, or “not noticeable” as they always noticed changes in the AV’s driving behavior. Thus, each participant interpreted the scale differently and the results showed a large deviation in answers (Fuest, Maier et al., 2020). As a result, the scale was changed and a five-point Likert scale was used as response format (chapter 9). The results showed that pedestrians perceived the situation as less critical when the AV yielded for them (section 12.3, Fuest, Michalowski et al., 2018). Using the item, it is also possible to differentiate between unambiguous and ambiguous driving profiles, as pedestrians perceived it as more critical if the AV drove ambiguously (Fuest, Schmidt, & Bengler, 2020). Thus, the results of the perceived criticality confirm the findings of the misinterpretation rate. However, the item depends on the personal circumstances. Discussions with the participants revealed that they included personal life situations in their evaluation, such as walking with their toddler. As a result, the intention “AV goes first” poses more potential risk for an accident, as compared to an AV yielding for the HRU. Therefore, the misinterpretation rate assesses the actual risk better, but leaves out the subjective perspective. Depending on the research question, it should be considered whether it is useful to evaluate the perceived criticality.

The results of the evaluation of driving behavior showed that the item can distinguish between different driving profiles as well (Fuest, Maier et al., 2020; Fuest, Michalowski et al., 2018; Fuest, Michalowski et al., 2019; Fuest, Schmidt, & Bengler, 2020). It was even possible to differentiate between driving profiles with low misinterpretation rates (Fuest, Michalowski et al., 2018). This makes it possible to select the better rated driving profile among different driving profiles with a low safety risk.

The subjective decision-making reliability is of limited help in distinguishing between different driving profiles. The participants were not in all four setups more confident in their decision when confronted with an unambiguous driving profile, as compared to an ambiguous driving profile. The added value of the results compared to the misinterpretation rate is therefore very low and the item can be omitted.

12.2.3 Transferability of the Metrics

The metrics used were developed to enable a distinction between different driving profiles that can be used to communicate a yielding or a non-yielding intention to pedestrians. In addition, the metrics were intended to evaluate to what extent the results of the different setups are comparable with each other. However, the IRT and the questionnaire items are not limited to investigate AV’s driving behavior.

IRT and questionnaire items can also be used to compare different eHMIs (e.g., Dietrich, Boos et al., 2020; Weber et al., 2019). It is even possible to analyze the IRT itself when comparing eHMIs. The prerequisite for the comparison is that the driving behavior will be kept constant and only the eHMI will be varied. First results have shown that with the help of the IRT differences can be found between different eHMIs in a VR study (Weber et al., 2019). The aim of the studies was to detect the most suitable eHMI. Therefore, the correct interpretation of the AV’s intention was used as a metric instead of the misinterpretation rate. Both, the IRT and the correct intention rate helped to differentiate between different eHMIs (Weber et al., 2019).

In addition, the metrics and the questionnaire items can be used to evaluate the driving behavior and eHMIs in other situations. Using the taxonomy of traffic situations it is possible to extend and adapt the situation (chapter 7; Fuest, Sorokin et al., 2018). Metrics and questionnaire items were already used in situations with different right of ways, AV’s intentions regarding the right of way, longitudinal and lateral distances, as well as the AV’s speed (chapter 8 to chapter 11). However, it is also possible to adjust the value facets of all other attributes.

12.3 Recommendations for the Driving Profile of Automated Vehicles

Different driving profiles were investigated for the intentions “AV goes first” (Table 12.1) and “Let the HRU go first” (Table 12.2) in the studies. Both intentions were communicated by using driving profiles that differed in longitudinal and, for the intention “AV goes first”, also in lateral dynamics.

Results revealed that pedestrians were able to understand the AV’s implicit communication (Fuest, Maier et al., 2020; Fuest, Michalowski et al., 2018; Fuest, Michalowski et al., 2019). This is in line with previous findings (Dey et al., 2019; Lundgren et al., 2017). For both intentions it can be deduced from study results that the AV should communicate the intention at a distance of 8 m to 25 m from the pedestrian’s position, as most pedestrians’ intention detections are made in this range (Fuest, Michalowski et al., 2019). Changing the intention during or after this range is not advisable and leads to a higher misinterpretation rate. In addition, ambiguous driving profiles might lead to more accidents because participants do not understand the AV’s intention (Fuest, Michalowski et al., 2019; Fuest, Schmidt, & Bengler, 2020).

12.3.1 Driving Profile for the Intention “AV goes first”

For the intention “AV goes first”, four driving profiles (Table 12.1) were evaluated with two different speeds: 49.0 km/h and 28.5 km/h (30 km/h indicated on the speedometer). During three driving profiles, the AV maintained speed, whereas during one profile the AV decelerated to 13 km/h and accelerated again at a 7.4 m distance from the pedestrian’s position. Furthermore the AV’s lateral position was varied: three driving profiles used a constant position, whereas in the fourth profile the AV drove 50 cm further towards the middle of the road.

The results for the evaluation of the driving behavior demonstrated that pedestrians preferred a speed of 28.5 km/h when the AV was driving in an undefined right of way situation over a speed of 49.0 km/h. A speed of 49.0 km/h was perceived as too fast in a parking area (chapter 8; Fuest, Maier et al., 2020). In addition, pedestrians favored a constant vehicle speed, rather than a speed reduction followed by an acceleration (Fuest, Michalowski et al., 2019; Fuest, Schmidt, & Bengler, 2020). The results regarding the lateral position did not show a clear tendency. In the direct comparison, the driving profile with the lateral offset towards the middle of the road was rated significantly better compared to a driving profile without lateral offset (Fuest, Michalowski et al., 2018). The driving profile with a speed of 28.5 km/h and without lateral offset was also used in three other studies with different setups (WoZ, VR, and videos; Fuest, Michalowski et al., 2018; Fuest, Schmidt, & Bengler, 2020). Compared to the results of these studies, the evaluation of the driving profile with the lateral offset was not better or worse. In addition, the misinterpretation rate was also not lower when the AV drove with a lateral offset. The direct comparison of the two WoZ studies also showed that the driving profile with the lateral offset was not generally rated better. However, the lateral offset increased the safety distance to the pedestrian and thus increased the possibility of emergency braking if the pedestrian steps onto the road by mistake. Nevertheless, this driving profile is only practicable if the road is wide enough and there is no oncoming traffic. As a result, the lateral offset would not be consistently drivable across all situations and could irritate pedestrians. Therefore, a recommendation for lane keeping was derived. Further evaluation is encouraged.

To sum up, the AV should maintain a speed that is adapted to the situation (here: 28.5 km/h) without a lateral offset to communicate a non-yielding intention.

Table 12.1 Driving profiles for the intention "AV goes first" (for undefined right of way situations).

Setup	Speed	Specification	Evaluation of Driving Behavior (1 = very poor, 5 = very good)	Misinterpretation Rate
VR[1]	49.0 km/h		2.0	1.6%
VR[3]	28.5 km/h		4.0	2.7%
Video VR[3]			4.0	2.4%
Video WoZ[3]			4.0	7.3%
WoZ[3]			4.5	3.2%
WoZ[2]			3.0	8.3%
VR[3]	28.5 km/h	the AV decelerated after 60 m to 13 km/h and accelerated again at a 7.4 m distance from the pedestrian's position	2.0	40.5%
Video VR[3]			2.0	27.9%
Video WoZ[3]			2.0	31.7%
WoZ[3]			2.0	71.0%
WoZ[2]	28.5 km/h	50 cm offset towards the middle of the road	4.0	6.9%

[1]Fuest, Maier et al. (2020), chapter 8; [2]Fuest, Michalowski et al. (2018), chapter 9; [3]Fuest, Michalowski et al. (2019), chapter 10 and Fuest, Schmidt, and Bengler (2020) chapter 11.

12.3.2 Driving Profile for the Intention "Let the HRU go first"

Five driving profiles with two different speeds—49.0 km/h and 28.5 km/h (30 km/h indicated on the speedometer)—were evaluated for the intention "Let the HRU go first" (Table 12.2). In all four driving profiles the AV decelerated constantly with different decelerations between -1.5 m/s^2 and -11.0 m/s^2. The beginning of the deceleration varied between 14.8 m and 40.0 m distance from the pedestrian's position and a complete stop between 4.3 m and 9.0 m from the pedestrian's position. In general, the results support the previous finding (Dey et al., 2019; Dietrich, Maruhn et al., 2020) that pedestrians understand a yielding intention based on the AV's braking behavior (Fuest, Maier et al., 2020; Fuest, Michalowski et al., 2018; Fuest, Michalowski et al., 2019).

The comparison of the driving profiles for the yielding intention is more complex because the driving profiles differed at various points and were not systematically varied. This was due to the fact that the main focus of this thesis is the comparison of methods and not the target design of driving profiles. However, when the right of way was undefined and the AV intended to yield the initial speed did not seem to have a strong influence on the evaluation and the misinterpretation rate (comparing the speed of 28.5 km/h and 49.0 km/h). In contrast, the start of deceleration had an influence. Braking later resulted in lower ratings. In addition, a complete stop at a 9.0 m distance from the pedestrian's position was rated better than a stop at a 4.3 m distance from the pedestrian's position. Driving with a speed of 28.5 km/h and starting to decelerate with at maximum -2.4 m/s^2 at a 22.0 m distance from the pedestrian's position—this is equivalent to a TTA (time to arrival) of 2.8 s—and a complete stop at a 9.0 m distance from the pedestrian's position was rated best when comparing all driving profiles. In addition, the misinterpretation rate was low for this driving profile. The worst rating and the highest misinterpretation rate were found for an abrupt and strong braking, i.e., decelerating with a maximum of -4.1 m/s^2 at a 14.8 m distance from the pedestrian's position (TTA = 1.9 s) and coming to a complete stop at a 4.3 m distance from the pedestrian's position. In general, it can be stated that pedestrians prefer AVs to yield for them (Fuest, Michalowski et al., 2018).

Summarizing, to communicate a yielding intention, the AV should decelerate with at maximum -2.4 m/s^2 starting at a 22.0 m distance from the pedestrian's position and come to a complete stop at a 9.0 m distance from the pedestrian's position when the initial speed is 30 km/h (indicated on the speedometer).

Table 12.2 Driving profiles for the intention "Let the HRU go first" (for undefined right of way situations).

Setup	**Speed**	**Max. Acceleration**	**Start of Acceleration** (distance from the pedestrian's position)	**Complete Stop** (distance from the pedestrian's position)	**Evaluation of Driving Behavior** (1 = very poor, 5 = very good)	**Misinterpre-tation Rate**
VR[1]	49.0 km/h	-6.0 m/s^2	25.0 m	5.0 m	4.0	4.9%
VR[1]	49.0 km/h	-11.0 m/s^2	16.0 m	5.0 m	3.0	29.9%
WoZ[2]	28.5 km/h	-2.4 m/s^2	22.0 m	9.0 m	5.0	3.6%
VR[3]	28.5 km/h	-1.5 m/s^2	40.0 m	7.4 m	4.5	2.7%
Video VR[3]					4.0	4.3%
Video WoZ[3]					4.0	11.1%
WoZ[3]					4.0	0.0%
VR[3]	28.5 km/h	-4.1 m/s^2	14.8 m	4.3 m	3.0	29.7%
Video VR[3]					2.0	63.6%
Video WoZ[3]					2.0	61.5%
WoZ[3]					3.5	77.4%

[1]Fuest, Maier et al. (2020), chapter 8; [2]Fuest, Michalowski et al. (2018), chapter 9; [3]Fuest, Michalowski et al. (2019), chapter 10 and Fuest, Schmidt, and Bengler (2020) chapter 11.

12.3.3 Transferability of the Driving Profiles

Comparing the investigated driving profiles with driving profiles from other studies is difficult for several reasons. Firstly, the driving parameters are not specified in most papers. Secondly, the study setups varied. As already assumed in chapter 3, confirmed in section 12.1, and shown in the column for the evaluation of driving behavior and misinterpretation rate in Table 12.1 and Table 12.2, results largely depend on the setup.

Since the results of the WoZ study are closest to reality, it is most useful to compare the presented results with other WoZ studies. However, there is very little research on driving behavior in WoZ setups. Currano et al. (2018) described that the driver was trained to “drive in a manner suggestive of how an automated system might drive: slowly and cautiously, in a consistent driving style“ (Currano et al., 2018, p. 213). However, driving data were not recorded. Similar statements can be made about the study by Rothenbücher et al. (2016). The driver was trained to either “accelerate in a safe way that would attract attention” (Rothenbücher et al., 2016, p. 798) or to drive “more aggressive and ambiguous” (Rothenbücher et al., 2016, p. 799). Specific driving parameters were not published and driving data were not recorded. In the experiment of Lundgren et al. (2017) the vehicle drove with only 7 km/h either without slowing down or the vehicle stopped about 3 m from the pedestrian’s position. The focus of the study was on the behavior of the driver, which is why the driving profiles were not further compared. Moore et al. (2019) used video data from the outside perspective of the vehicle to analyze at what distance the vehicle stopped in front of a crosswalk. Distances from 1 m to 6 m were used. Results showed that pedestrians understand the yielding intention better, when the vehicle came to a standstill closer to the crosswalk. Pedestrians perceived a standstill at a 2 m distance from the crosswalk as most clear (Moore et al., 2019). The authors did not specify when the vehicle started braking.

As the results show, the published driving parameters are not sufficient to derive conclusions or to compare data. Transferring the driving data published in the thesis to other situations is similarly difficult. This is partly due to the fact that a very reduced and unnatural situation was used for the method comparison. One-to-one situations—one pedestrian and one AV—rarely exist in reality with the pedestrian waiting at an initial distance of 100 m until the vehicle starts to decelerate. However, referring to the taxonomy from chapter 7 (Fuest, Sorokin et al., 2018), it can be assumed that the driving data can be applied to similar situations as shown in Table 7.1 and by the adaptations described in chapter 8. The driving data can be transferred to situations in which the AV has an initial speed of 30 km/h and communicates a yielding or non-yielding intention to an undistracted pedestrian standing in the AV’s field of view at the roadside in the direction of travel (see Fuest, Sorokin et al., 2018). The results of the study in chapter 8 have shown that the driving profile is independent of the right of way (Fuest, Maier et al., 2020). Therefore, the driving data can be transferred to situations in which the AV has the right of way, the HRU has the right of way, or the right of way situation is not defined. It also does not matter whether a passenger is sitting on the driver's seat or whether it is empty (Fuest, Michalowski et al., 2018). As already shown, the results are transferable to situations where the AV has an initial speed of 49 km/h (Fuest, Maier et al., 2020). For higher speeds outside urban areas, the start of braking must be adjusted depending on the initial speed so that the vehicle can come to a standstill in time. At lower speeds a later braking and a later standstill also seem to be accepted by pedestrians (Lundgren et al., 2017).

The following assumptions can be made for the transferability of the driving data to situations with other value facets, as long as only one value facet is adjusted and all others are kept constant. Since the pedestrian was always facing the AV, it can be assumed that the driving profile can be used regardless of the approach direction (right or left of the pedestrian). This only applies as long as no other vehicle is coming from the other direction. A transfer of the driving profile to situations in which communication with other HRUs, instead of a pedestrian, occurs might be possible given that all other attributes are kept constant. As long as other HRUs, such as drivers and cyclists, are standing still during the AV's communication, the perspective and the perception are similar. Only the crossing behavior would presumably change, since particularly drivers can accelerate faster than pedestrians. However, it can be assumed that the results cannot be transferred to situations in which the HRU is moving. Communication must take place later and therefore faster as the AV is only perceived later by the HRU. The same applies if there is an impairment of the HRU's perception, for instance, parked vehicles on the roadside. For such situations, it needs to be investigated as to what extent the driving profile can be perceived by the sound of the AV's engine and roll noises. However, this problem already exists in today's road traffic, for instance when pedestrians stand between parked vehicles before crossing the road and cannot see approaching vehicles. The same applies to distracted pedestrians such as those using smartphones. In such situations, pedestrians cannot see the driving profile of the approaching AV. More research is needed whether eHMIs, for example with sound, might be beneficial in these situations. The situation is different when pedestrians are distracted by music through headphones. However, transferability is also not given when the driving profile is not viewed from a sideways perspective but frontally, for example when the AV has to communicate with another vehicle in a bottleneck. In addition, it is not possible to transfer the driving data to smaller longitudinal distances because in the published studies the AV starts decelerating at a greater distance. Further studies have to be conducted for smaller distances and possibly lower speeds, as otherwise the braking distance would be too large (at higher speeds) for the AV to yield.

13 Future Work

The comparison of the methods (section 12.1), the description of the metrics used (section 12.2), as well as the description and the transferability of different driving profiles (section 12.3) demonstrate that there are still open questions regarding the research of evaluating AV’s driving behavior.

Regarding the comparison of methods, it remains to be verified to what extent the use of an experimental vehicle provides advantages as compared to a WoZ vehicle. However, there are also additional questions regarding the WoZ vehicle. For example, the reproducibility should be further examined. It is necessary to evaluate more precisely at which deviations differences in driving profile are perceivable for participants and at which point these deviations lead to different results that are relevant for traffic safety. This does not only apply to changes in longitudinal driving parameters, but also to a lateral offset. The latter has not yet been investigated at all. As mentioned in section 12.1.4, a limitation of the method comparison consists in the differences for view and sound of participants. While in the VR and video setups the situation froze, participants turned to face the experiencer in the WoZ setup. A possible solution might be the use of shutter glasses and a noise cancelling headphone. At the moment participants recognize the intention and press the button, shutter glasses and headphone should suppress view and hearing.

In addition, for the VR and video setup, it needs to be examined whether and to what extent engine sounds could improve the results. It can be assumed that VR technology is constantly improving and perception is getting increasingly better. As a result, the disadvantages of the VR setup that were pointed out in section 12.1 could become fewer with technological development. This needs to be further reviewed.

Regarding objective metrics, the IRT as well as all other metrics analyzing a time response to different driving profiles are not independent of the driving profiles themselves (subsection 12.2.1). Here, a metric needs to be developed that analyzes the response of pedestrians to the AV’s driving behavior regardless of the duration of the driving profile. Since the focus of the thesis was on the comparison of methods and the use of the devised metrics, different metrics were not systematically compared. Although the crossing behavior was compared with the IRT in the VR setup, a further comparison with other metrics (chapter 4) is needed.

The same applies to the derivation of suitable driving profiles for AVs. Driving profiles based on human driving behavior were used for the method comparison. However, a systematic comparison is lacking, especially in order to be able to ensure transferability to other situations (subsection 12.3.3). It is especially important to investigate more realistic situations with multiple road users. One question would be whether pedestrians can recognize for whom the vehicle is braking when there are several pedestrians at different distances at the roadside. Another question is whether the driving behavior is also sufficient when there is less distance to communicate the intentions, because another vehicle is driving in front of the AV and blocks the view.

As the results of the study comparison (chapter 11; Fuest, Schmidt, & Bengler, 2020) showed differences in the variances between the driving profiles and setups, the reliability should be further examined. Furthermore, the methods need to be validated for the specific use case, as this was not the goal of the thesis.

Referring to the *Sender Receiver Model* (chapter 2), it can be stated that it is possible that an AV's intention can be communicated by encoding the message through implicit communication. When an unambiguous driving profile is used, a pedestrian can decode the message correctly. However, it can be assumed that different environmental influences, such as a visual impairment and weather conditions, e.g., fog or rain, generate noise. Beggiato et al. (2018) showed that dusk interferes with the transmission and leads to more conservative time gaps, as compared to a day condition. However, open questions remain, as Ackermann, Beggiato, Bluhm et al. (2019) could not verify this effect. For method comparison, metric development, and the evaluation of the driving behavior, the *Sender Receiver Model* has been shown to be sufficient as a first step. However, as described in chapter 2, traffic behavior includes a social component (Klebelsberg, 1982). This component was illustrated in this research, since the item regarding the perceived criticality depends on personal circumstances (subsection 12.2.2). Therefore, an extended model of communication should be used in a next step. The *Four-Sides Model* (Schulz von Thun, 2018) can be recommended, as this model includes, besides the factual information, a self-revelation, a relationship between sender and receiver, and an appeal (Schulz von Thun, 2018). It has to be decided whether the AV communicates its driving status (manual vs. automated) at the self-revelation level. At the relationship level, the AV uses the information gathered about the particular HRU with whom communication is taking place. At the appeal level, it must be ensured that the AV does not communicate an appeal, but only its intention (Andersson et al., 2017; Zhang et al., 2018). Depending on the research question, the four levels must also be considered for the pedestrian. Thereby, e.g., pedestrian's trust in and acceptance of the AV and the motivation should be taken into account.

These open questions need to be addressed before AVs can be integrated into road traffic. Otherwise, ambiguous driving profiles could lead to accidents and thus decreased acceptance of AVs.

14 Conclusion

Regarding the research question “What method can be recommended to evaluate the AV's driving behavior?” it can be stated that it is possible to detect ambiguous driving profiles in all setups—Woz, VR and videos—by using the misinterpretation rate and the mentioned hypothetical crossing behavior. A VR setup can be used to preselect ambiguous driving profiles. However, since the results partly underestimate the misinterpretation rate and overestimate the potential of driving profiles, further evaluation of the selected driving profiles in a WoZ setup should follow. Since the effort of a video study is not in proportion to the limited results, the use of video studies for the evaluation of driving profiles is not recommended. Instead, the use of a WoZ approach is recommended to evaluate different driving profiles. However, one weakness of the WoZ approach is the lack of reproducibility of the target driving profile. This factor needs to be considered when choosing the driving profiles to be evaluated. Consequently, driving data has to be recorded during a WoZ study. However, a quality standard that must be fulfilled for all setups is the precise specification of the driving profile.

To answer the research question “Can the metrics be recommended to distinguish between different driving profiles?” the following advice can be given. The IRT should not be evaluated separately, when comparing different driving profiles, as reaction times always depend on the duration of the driving profile. In addition, the percentage of those participants who waited to press the button until the AV came to a complete standstill or passed by cannot be recommended to differentiate between unambiguous and ambiguous driving profiles. Results showed that it was independent of the unambiguity of the driving profiles that pedestrians made their decision before the AV stopped completely or passed by. However, the moment participants recognized the AV’s intention is an appropriate time for asking additional questionnaire items. Participants should be asked to rate the AV’s driving behavior to differentiate between two unambiguous driving profiles. This can be used to ensure that a driving profile is not only unambiguous and safe for pedestrians, but is also accepted. In addition, it is recommended to enquire about the assumed intention of the AV to analyze the misinterpretation rate. This allows for detecting ambiguous and safety-critical driving profiles. The same applies to the mentioned hypothetical crossing behavior to analyze the crossing rate. This metric cannot only identify risky driving profiles, but also discover the potential of unambiguous driving profiles. The two metrics eliminate the need to ask participants about perceived criticality. Moreover, it is not necessary to let participants rate their certainty about the AVs intention.

Regarding the last research question “What driving behavior can be recommended for AVs to communicate a yielding or a non-yielding intention to a pedestrian?” it can be deduced from the results that the AV should communicate its intention at a distance of 8 m to 25 m from the pedestrian’s position when the initial speed is 30 km/h (indicated on the speedometer). For a non-yielding intention, the AV should maintain speed. When the AV indents to yield, it should decelerate with at maximum -2.4 m/s^2 starting at a 22.0 m distance and a complete stop at a 9.0 m distance from the pedestrian’s position.

15 Complementary Studies and Contributor Roles

In addition to the thesis, further studies have been conducted that were not in the main focus of the thesis but nevertheless contribute to the communication between AVs and HRUs. Three of the studies focus on the use of explicit communication, and one on the abstraction level of the environment.

Weber, F., Chadowitz, R., Schmidt, K., Messerschmidt, J., & Fuest, T. (2019). Crossing the street across the globe: a study on the effects of eHMI on pedestrians in the US, germany and china. In *21st International Conference on Human-Computer Interaction*, Florida, USA, 26.-31.07.2019.

The driving profiles from the first VR setup (chapter 8) and metrics were used to evaluate two different eHMIs in a VR setup—a light band and a display with icon—in three different countries—United States, Germany and China—regarding their impact on pedestrians. Instead of the misinterpretation rate, the correct interpretation rate was analyzed, as the aim was not to identify the worst eHMI, but the best one. Results revealed a positive impact on the correct interpretation rate for the light band in Germany and the United States, when the AV intended to yield. In contrast, no impact was found for the display. The eHMI showed no positive impact in China. In addition, eHMIs deteriorated the correct interpretation rate when the AV did not yield across all countries. It can be concluded that eHMIs are not helpful in all conditions and situations in which an eHMI improves the communication should be selected carefully. Additionally, cultural adaptions have to be evaluated before integrating an eHMI internationally.

Song, Y. E., Lehsing, C., Fuest, T., & Bengler, K. (2018). External HMIs and Their Effect on the Interaction Between Pedestrians and Automated Vehicles. In W. Karwowski & T. Ahram (Eds.), *Advances in Intelligent Systems and Computing, Intelligent Human Systems Integration* (pp. 13–18). Springer International Publishing.

The study focused on the influence of different eHMIs on the pedestrians' crossing behavior. Two eHMIs—one with a command message (Go!) and one with an affirmative message (Ok!)—were compared with a baseline condition—showing no eHMI—in a video setup. Pedestrians had to decide whether they would cross (by pressing the space key on a keyboard) or not cross the road (by pressing no key). Results revealed that pedestrians crossed the road more often when the AV was equipped with an eHMI. However, no differences were found between the two eHMI message types.

Fuest, T., Feierle, A., Schmidt, E., & Bengler, K. (2020). Effects of Marking Automated Vehicles on Human Drivers on Highways. *Information*, 11 (6), 286.

It can be assumed that when integrating AVs in mixed traffic the initial driving behavior may differ from manual road traffic. Therefore, marking the AVs as driving automatically might be advantageous. To validate this hypothesis, a simulator study was conducted. In three different highway situations, drivers encountered an AV driving very conservatively and strictly adhering to German highway regulations. The vehicle was either marked as driving automatically or drove without a marking. Driving data were measured and the time of lane change and the time headway were analyzed. In addition, participants had to rate how disturbing and appropriate they perceived the AV's driving behavior. Results revealed no differences between a marked and a non-marked AV. It can be assumed that drivers have compensation strategies to deal with vehicles that drive differently. In addition, it seems that the AV's driving behavior is sufficient to recognize the vehicle as automated. Nevertheless, when asking participants directly they preferred a marking in order to distinguish between AVs and other vehicles.

Angerstorfer, J., dem Bruch, L. aus, Nourian, S., & Tannert, I. (2019). *Einfluss der Abstraktion von Videos auf die Erkennung der Intention automatisierter Fahrzeuge durch Fußgänger* [Bericht zum Interdisziplinären Projekt; Betreuung durch T. Fuest]. Technische Universität München.

During this student project, it was evaluated how detailed a scene has to be so that a pedestrian is able to recognize an AV's intention. Therefore, the abstraction level was varied. In the first condition, only the vehicle was presented to the participants. In the second condition, houses, trees and a road were added to the scene for a lower abstraction level. In the study, the AV was represented by an abstract cuboid. In addition, it was analyzed whether pedestrians are able to recognize the AV's intention, when they saw the driving behavior from a bird's eye view, as used in driving school questionnaires. The driving profiles used in the second WoZ study and in the method comparison (chapter 10 and chapter 11) were implemented in VR and videos were recorded for a video setup. The dependent variables presented in the thesis were used to analyze the influence of the abstraction level and the perspective. The results revealed no differences for the abstraction level. However, using the bird's eye view intentions were more often interpreted correctly. This could be due to the fact that participants in the bird's-eye view need longer to recognize the AV's intention. Due to the extended IRT, participants observed the driving behavior over a longer period of time and were therefore able to better assess the intention of the vehicle.

Contributor Roles to the Publications for the Thesis

Fuest, T., Sorokin, L., Bellem, H., & Bengler, K. (2018). Taxonomy of Traffic Situations for the Interaction between Automated Vehicles and Human Road Users. In N. A. Stanton (Ed.), *Advances in Intelligent Systems and Computing. Advances in Human Aspects of Transportation* (Vol. 597, pp. 708–719). Cham: Springer International Publishing.

Conceptualization: T.F., L.S., and H.B.; Supervision: T.F., L.S., and K.B.; Visualization: T.F.; Writing—original draft: T.F.; Writing—review and editing: T.F., L.S., H.B., and K.B.

Fuest, T., Maier, A.S., Bellem, H., Bengler, K. (2020). How Should an Automated Vehicle Communicate Its Intention to a Pedestrian? – A Virtual Reality Study. In: Ahram T., Karwowski W., Pickl S., Taiar R. (eds) *Human Systems Engineering and Design II. IHSED 2019. Advances in Intelligent Systems and Computing*, Vol 1026. Springer, Cham.

Conceptualization: T.F., A.S.M., and H.B.; Data curation: T.F. and A.S.M.; Formal analysis: T.F.; Investigation: T.F. and A.S.M.; Methodology: T.F.; Supervision: T.F., H.B., and K.B.; Visualization: T.F.; Writing—original draft: T.F.; Writing—review and editing: T.F., H.B., and K.B.

Fuest, T., Michalowski, L., Träris, L., Bellem, H., & Bengler, K. (2018). Using the Driving Behavior of an Automated Vehicle to Communicate Intentions – A Wizard of Oz Study. In *21st International Conference on Intelligent Transportation Systems* (ITSC) (pp. 3596–3601). IEEE.

Conceptualization: T.F., L.M., and L.T.; Data curation: T.F., L.M., and L.T.; Formal analysis: T.F.; Investigation: T.F., L.M., and L.T.; Methodology: T.F.; Supervision: T.F., H.B., and K.B.; Visualization: T.F.; Writing—original draft: T.F.; Writing—review and editing: T.F., L.M., H.B., and K.B.

Fuest, T., Michalowski, L., Schmidt, E., & Bengler, K. (2019). Reproducibility of Driving Profiles–Application of the Wizard of Oz Method for Vehicle Pedestrian Interaction. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)* (pp. 3954-3959). IEEE.

Conceptualization: T.F. and L.M.; Data curation: T.F., and L.M.; Formal analysis: T.F. and L.M.; Investigation: T.F., and L.M.; Methodology: T.F.; Supervision: T.F. and K.B.; Visualization: T.F., L.M., and E.S.; Writing—original draft: T.F.; Writing—review and editing: T.F., L.M., E.S., and K.B.

Fuest, T., Schmidt. E., & Bengler, K. (2020). Comparison of Methods to Evaluate the Influence of Automated Vehicle's Driving Behavior on Pedestrians: Wizard of Oz, Virtual Reality, and Video. *Information* 11 (6), 291.

Conceptualization: T.F., E.S, and K.B.; Data curation: T.F..; Formal analysis: T.F.; Investigation: T.F.; Methodology: T.F.; Supervision: T.F., and K.B.; Visualization: T.F. and E.S.; Writing—original draft: T.F.; Writing—review and editing: T.F., E.S., and K.B.

16 References

AASHTO. (2011). *A Policy on geometric design of highways and streets* (6th ed.).

Ackermann, C., Beggiato, M., Bluhm, L.-F., & Krems, J. F. (2018). Vehicle Movement and its Potential as Implicit Communication Signal for Pedestrians and Automated Vehicles. In N. Van Nes & C. Voegelé (Eds.), *Proceedings of the 6th HUMMANIST Conference.* HUMANIST publications. https://www.humanist-vce.eu/fileadmin/contributeurs/humanist/TheHague2018/Proceedings.pdf

Ackermann, C., Beggiato, M., Bluhm, L.-F., Löw, A., & Krems, J. F. (2019). Deceleration parameters and their applicability as informal communication signal between pedestrians and automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *62*, 757–768. https://doi.org/10.1016/j.trf.2019.03.006

Ackermann, C., Beggiato, M., Schubert, S., & Krems, J. F. (2019). An experimental study to investigate design and assessment criteria: What is important for communication between pedestrians and automated vehicles? *Applied Ergonomics*, *75*, 272–282. https://doi.org/10.1016/j.apergo.2018.11.002

Allen, R. W., Rosenthal, T. J., & Cook, M. L. (2011). A Short History of Driving Simulation. In D. L. Fisher, M. Rizzo, J. K. Caird, & J. D. Lee (Eds.), *Handbook of driving simulation for engineering, medicine, and psychology* (2.1-2.16). CRC Press.

Andersson, J., Habibovic, A., Klingegård, M., Englund, C., & Malmsten-Lundgren, V. (2017). *Hello Human, can you read my mind?* https://ercim-news.ercim.eu/en109/special/hello-human-can-you-read-my-mind

Angerstorfer, J., dem Bruch, L. aus, Nourian, S., & Tannert, I. (2019). *Einfluss der Abstraktion von Videos auf die Erkennung der Intention automatisierter Fahrzeuge durch Fußgänger* [Bericht zum Interdisziplinären Projekt]. Technische Universität München, München.

Beggiato, M., Witzlack, C., & Krems, J. F. (2017). Gap Acceptance and Time-To-Arrival Estimates as Basis for Informal Communication between Pedestrians and Vehicles. In S. Boll, B. Pfleging, I. Politis, D. Large, & B. Domnez (Eds.), *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '17* (pp. 50–57). ACM Press. https://doi.org/10.1145/3122986.3122995

Beggiato, M., Witzlack, C., Springer, S., & Krems, J. (2018). The Right Moment for Braking as Informal Communication Signal Between Automated Vehicles and Pedestrians in Crossing Situations. In N. A. Stanton (Ed.), *Advances in Intelligent Systems and Computing. Advances in Human Aspects of Transportation* (Vol. 597, pp. 1072–1081). Springer International Publishing. https://doi.org/10.1007/978-3-319-60441-1_101

Bengler, K. (2015). Messmethoden: Notwendigkeit von Versuchen. In H. Bubb, K. Bengler, R. E. Grünen, & M. Vollrath (Eds.), *ATZ / MTZ-Fachbuch. Automobilergonomie* (pp. 618–621). Springer Vieweg.

Bengler, K., Omozik, K., & Müller, A. I. (2020). The Renaissance of Wizard of Oz (WoOz)– Using the WoOz methodology to prototype automated vehicles. In D. de Waard, A. Toffetti, L. Pietrantoni, T. Franke, J.-F. Petiot, C. Dumas, A. Botzer, L. Onnasch, I. Milleville, & F. Mars (Eds.), *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2019 Annual Conference* (pp. 63–72). ISSN 2333-4959.

Bengler, K., Rettenmaier, M., Fritz, N., & Feierle, A. (2020). From HMI to HMIs: Towards an HMI Framework for Automated Driving. *Information*, *11*(2), 61. https://doi.org/10.3390/info11020061

Bhagavathula, R., Williams, B., Owens, J., & Gibbons, R. (2018). The Reality of Virtual Reality: A Comparison of Pedestrian Behavior in Real and Virtual Environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 2056–2060. https://doi.org/10.1177/1541931218621464

Böckle, M.-P., Brenden, A. P., Klingegård, M., Habibovic, A., & Bout, M. (2017). SAV2P - Exploring the Impact of an Interface for Shared Automated Vehicles on Pedestrians' Experience. In A. Löcken, S. Boll, I. Politis, S. Osswald, R. Schroeter, D. Large, M. Baumann, I. Alvarez, L. Chuang, S. Feuerstack, M. Jeon, H. H. van Huysduynen, & N. Broy (Eds.), *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct - AutomotiveUI '17* (pp. 136–140). ACM Press. https://doi.org/10.1145/3131726.3131765

Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G [G.], & Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 708–713). IEEE. https://doi.org/10.1109/IROS.2005.1545011

Brown, B., & Laurier, E. (2017). The Trouble with Autopilots: Assisted and Autonomous Driving on the Social Road. In G. Mark, S. Fussell, C. Lampe, M. C. Schraefel, J. P. Hourcade, C. Appert, & D. Wigdor (Eds.), *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17* (pp. 416–429). ACM Press. https://doi.org/10.1145/3025453.3025462

Bundesamt für Justiz. (2013). *Straßenverkehrs-Ordnung (StVO)*. https://www.gesetze-im-internet.de/stvo_2013/StVO.pdf

Burns, C. G., Oliveira, L., Hung, V., Thomas, P., & Birrell, S. (2020). Pedestrian Attitudes to Shared-Space Interactions with Autonomous Vehicles – A Virtual Reality Study. In N. Stanton (Ed.), *Advances in Intelligent Systems and Computing, Advances in Human Factors of Transportation* (pp. 307–316). Springer International Publishing. https://doi.org/10.1007/978-3-030-20503-4_29

Burns, C. G., Oliveira, L., Thomas, P., Iyer, S., & Birrell, S. (2019). Pedestrian Decision-Making Responses to External Human-Machine Interface Designs for Autonomous Vehicles. In *2019 IEEE Intelligent Vehicles Symposium (IV)* (pp. 70–75). IEEE. https://doi.org/10.1109/IVS.2019.8814030

Cacilo, A., Schmidt, S [Sarah], Wittlinger, P., Herrmann, F., Bauer, W., Sawade, O., Doderer, H., Hartwig, M., & Scholz, V. (2015). *Hochautomatisiertes Fahren auf Autobahnen - Industriepolitische Schlussfolgerungen: Dienstleistungsprojekt 15/14 - Studie im Auftrag des Bundesministeriums für Wirtschaft und Energie (BMWi)*. https://www.bmwi.de/Redaktion/DE/Downloads/H/hochautomatisiertes-fahren-auf-autobahnen.pdf%3F__blob%3DpublicationFile%26v%3D1

Card, S. K., Moran, T. P., & Newell, A. (2008). *The Psychology of Human-Computer Interaction* (Reprinted.). CRC Press.

Castelfranchi, C., Pezzulo, G., & Tummolini, L. (2010). Behavioral Implicit Communication (BIC). *International Journal of Ambient Computing and Intelligence*, *2*(1), 1–12. https://doi.org/10.4018/jaci.2010010101

Chandler, D., & Munday, R. (2011). *A dictionary of media and communication*. Oxford Univ. Press.

http://www.oxfordreference.com/view/10.1093/acref/9780199568758.001.0001/acref-9780199568758

Chang, C.-M., Toda, K., Igarashi, T., Miyata, M., & Kobayashi, Y. (2018). A Video-based Study Comparing Communication Modalities between an Autonomous Car and a Pedestrian. In Unknown (Ed.), *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '18* (pp. 104–109). ACM Press. https://doi.org/10.1145/3239092.3265950

Chang, C.-M., Toda, K., Sakamoto, D., & Igarashi, T. (2017). Eyes on a Car: an Interface Design for Communication between an Autonomous Car and a Pedestrian. In S. Boll, B. Pfleging, I. Politis, D. Large, & B. Domnez (Eds.), *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '17* (pp. 65–73). ACM Press. https://doi.org/10.1145/3122986.3122989

Clamann, M., Aubert, M., & Cummings, M. L. (2017). Evaluation of Vehicle-to-Pedestrian Communication Displays for Autonomous Vehicles. In *Proceedings of the 96th Annual Transportation Research Board Meeting.* Symposium conducted at the meeting of Transportation Research Board (TRB), Washington, DC, United States.

Clausen, U., & Klingner, M. (2018). Automatisiertes Fahren: Computer greifen zum Steuer. In R. Neugebauer (Ed.), *Fraunhofer-Forschungsfokus. Digitalisierung: Schlüsseltechnologien für Wirtschaft und Gesellschaft* (1st ed., pp. 385–411). Springer Vieweg.

Clercq, K. de, Dietrich, A., Núñez Velasco, J. P., Winter, J. de [Joost], & Happee, R. (2019). External Human-Machine Interfaces on Automated Vehicles: Effects on Pedestrian Crossing Decisions. *Human Factors*, *61*(8), 1353–1370. https://doi.org/10.1177/0018720819836343

Cœugnet, S., Cahour, B., & Kraïem, S. (2019). A Psycho-Ergonomic Approach of the Street-Crossing Decision-Making: Toward Pedestrians' Interactions with Automated Vehicles. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Advances in Intelligent Systems and Computing. Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)* (Vol. 823, pp. 132–141). Springer International Publishing. https://doi.org/10.1007/978-3-319-96074-6_14

Currano, R., Park, S. Y., Domingo, L., Garcia-Mancilla, J., Santana-Mancilla, P. C., Gonzalez, V. M., & Ju, W. (2018). ¡Vamos! Observations of Pedestrian Interactions with Driverless Cars in Mexico. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '18* (pp. 210–220). ACM Press. https://doi.org/10.1145/3239060.3241680

Daimler AG. (2020). *Pioneering achievement: Autonomous long-distance drive in rural and: Mercedes-Benz S-Class INTELLIGENT DRIVE drives autonomously in the tracks of Bertha Benz*. https://media.daimler.com/marsMediaSite/en/instance/ko.xhtml?oid=9904223

Deb, S., Carruth, D. W., Sween, R., Strawderman, L., & Garrison, T. M. (2017). Efficacy of virtual reality in pedestrian safety research. *Applied Ergonomics*, *65*, 449–460. https://doi.org/10.1016/j.apergo.2017.03.007

Deb, S., Rahman, M. M., Strawderman, L. J., & Garrison, T. M. (2018). Pedestrians' Receptivity Toward Fully Automated Vehicles: Research Review and Roadmap for Future Research. *IEEE Transactions on Human-Machine Systems*, 1–12.

Deb, S., Strawderman, L. J., & Carruth, D. W. (2018). Investigating pedestrian suggestions for external features on fully autonomous vehicles: A virtual reality experiment.

Transportation Research Part F: Traffic Psychology and Behaviour, *59*, 135–149. https://doi.org/10.1016/j.trf.2018.08.016

Dey, D., Martens, M., Eggen, B., & Terken, J. (2017). The Impact of Vehicle Appearance and Vehicle Behavior on Pedestrian Interaction with Autonomous Vehicles. In A. Löcken, S. Boll, I. Politis, S. Osswald, R. Schroeter, D. Large, M. Baumann, I. Alvarez, L. Chuang, S. Feuerstack, M. Jeon, H. H. van Huysduynen, & N. Broy (Eds.), *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct - AutomotiveUI '17* (pp. 158–162). ACM Press. https://doi.org/10.1145/3131726.3131750

Dey, D., Martens, M., Eggen, B., & Terken, J. (2019). Pedestrian road-crossing willingness as a function of vehicle automation, external appearance, and driving behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 191–205. https://doi.org/10.1016/j.trf.2019.07.027

Dey, D., & Terken, J. (2017). Pedestrian Interaction with Vehicles: Roles of Explicit and Implicit Communication. In S. Boll, B. Pfleging, I. Politis, D. Large, & B. Domnez (Eds.), *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '17* (pp. 109–113). ACM Press. https://doi.org/10.1145/3122986.3123009

Dietrich, A., Bengler, K., Markkula, G., Giles, O., Lee, Y. M., Pekkanen, J., Madigan, R., & Merat, N. (2019). *interACT D.2.2. Final description of psychological models on human-human and human-automation interaction*. https://www.interact-roadautomation.eu/wp-content/uploads/interACT_D2.2_Interaction_Models_190902_v1.0_website-3.pdf

Dietrich, A., Boos, A., Bengler, K., Lee, Y. M., Madigan, R., Merat, N., Weber, F., Tango, F., Avsar, H., Utesh, F., Portouli, E., & Nathanael, D. (September 2020). *interACT D.6.2. Evaluation report on on-board user and road users interaction with AVs equipped with the interACT technologies*. https://www.interact-roadautomation.eu/wp-content/uploads/interACT_D6_2_v1.0_FinalWebsite.pdf

Dietrich, A., Maruhn, P., Schwarze, L., & Bengler, K. (2020). Implicit Communication of Automated Vehicles in Urban Scenarios: Effects of Pitch and Deceleration on Pedestrian Crossing Behavior. In T. Ahram, W. Karwowski, S. Pickl, & R. Taiar (Eds.), *Advances in Intelligent Systems and Computing. Human Systems Engineering and Design II* (Vol. 1026, pp. 176–181). Springer International Publishing. https://doi.org/10.1007/978-3-030-27928-8_27

Dietrich, A., Willrodt, J.-H., Wagner, K., & Bengler, K. (2018). Projection-Based External Human Machine Interfaces–Enabling Interaction between Automated Vehicles and Pedestrians. In *Driving Simulation & Virtual Reality Conference & Exhibition,* Juan-les-Pins, Antibe, France.

Dönch, L. D. (2018). *Einflüsse unterschiedlicher Detailstufen der Virtuellen Realität in einem Fußgängersimulator* [Bachelorarbeit]. Technische Universität München, München.

Eisma, Y. B., van Bergen, S., ter Brake, S. M., Hensen, M. T. T., Tempelaar, W. J., & Winter, J. C. F. de (2020). External Human–Machine Interfaces: The Effect of Display Location on Crossing Intentions and Eye Movements. *Information*, *11*(1), 13. https://doi.org/10.3390/info11010013

Faas, S. M., & Baumann, M. (2020). Yielding Light Signal Evaluation for Self-driving Vehicle and Pedestrian Interaction. In T. Ahram, W. Karwowski, S. Pickl, & R. Taiar

(Eds.), *Human Systems Engineering and Design II* (Vol. 1026, pp. 189–194). Springer International Publishing. https://doi.org/10.1007/978-3-030-27928-8_29

Faas, S. M., Mathis, L.-A., & Baumann, M. (2020). External HMI for self-driving vehicles: Which information shall be displayed? *Transportation Research Part F: Traffic Psychology and Behaviour*, *68*, 171–186. https://doi.org/10.1016/j.trf.2019.12.009

Färber, B. (2016). Communication and Communication Problems Between Autonomous Vehicles and Human Drivers. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.), *Autonomous Driving* (Vol. 17, pp. 127–146). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-48847-8_7

Farmer, T. (1998). Using the internet for primary research data collection. *Journal of Computer-Mediated Communication*, *5*(3), 124–132.

Federal Ministry for Digital and Economic Affairs. (2020). *Verordnung des Bundesministers für Verkehr, Innovation und Technologie über Rahmenbedingungen für automatisiertes Fahren (Automatisiertes Fahren Verordnung – AutomatFahrV).* https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20009740

Fraser, N. M., & Gilbert, G.N. (1991). Simulating speech systems. *Computer Speech & Language*, *5*(1), 81–99. https://doi.org/10.1016/0885-2308(91)90019-M

Fridman, L., Mehler, B., Xia, L., Yang, Y., Facusse, L. Y., & Reimer, B. (2019). To Walk or Not to Walk: Crowdsourced Assessment of External Vehicle-to-Pedestrian Displays. In *Transportation Research Board Annual Meeting,* Washington, DC. http://arxiv.org/pdf/1707.02698v1

Fuest, T. (2019, February 8). *Außenwirkung automatisiert fahrender Fahrzeuge.* Technische Universität München. Lehrstuhl für Ergonomie. Lehrstuhlseminar, München.

Fuest, T., Feierle, A., Schmidt, E., & Bengler, K. (2020). Effects of Marking Automated Vehicles on Human Drivers on Highways. *Information*, *11 (6), 286.* https://doi.org/10.3390/info11060286

Fuest, T., Maier, A. S., Bellem, H., & Bengler, K. (2020). How Should an Automated Vehicle Communicate Its Intention to a Pedestrian? – A Virtual Reality Study. In T. Ahram, W. Karwowski, S. Pickl, & R. Taiar (Eds.), *Advances in Intelligent Systems and Computing. Human Systems Engineering and Design II* (Vol. 1026, pp. 195–201). Springer International Publishing. https://doi.org/10.1007/978-3-030-27928-8_30

Fuest, T., Michalowski, L., Schmidt, E., & Bengler, K. (2019, October). Reproducibility of Driving Profiles – Application of the Wizard of Oz Method for Vehicle Pedestrian Interaction. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)* (pp. 3954–3959). IEEE. https://doi.org/10.1109/ITSC.2019.8917395

Fuest, T., Michalowski, L., Träris, L., Bellem, H., & Bengler, K. (2018, November). Using the Driving Behavior of an Automated Vehicle to Communicate Intentions - A Wizard of Oz Study. In *21st International Conference on Intelligent Transportation Systems (ITSC)* (pp. 3596–3601). IEEE. https://doi.org/10.1109/ITSC.2018.8569486

Fuest, T., Schmidt, E., & Bengler, K. (2020). Comparison of Methods to Evaluate the Influence of an Automated Vehicle's Driving Behavior on Pedestrians: Wizard of Oz, Virtual Reality, and Video. *Information*, *11 (6), 291.* https://doi.org/10.3390/info11060291

Fuest, T., Sorokin, L., Bellem, H., & Bengler, K. (2018). Taxonomy of Traffic Situations for the Interaction between Automated Vehicles and Human Road Users. In N. A. Stanton

(Ed.), *Advances in Intelligent Systems and Computing. Advances in Human Aspects of Transportation* (Vol. 597, pp. 708–719). Springer International Publishing. https://doi.org/10.1007/978-3-319-60441-1_68

Gildert, N., Millard, A. G., Pomfret, A., & Timmis, J. (2018). The Need for Combining Implicit and Explicit Communication in Cooperative Robotic Systems. *Frontiers in Robotics and AI*, *5*, 13. https://doi.org/10.3389/frobt.2018.00065

Greenberg, J., & Blommer, M. (2011). Physical Fidelity of Driving Simulators. In D. L. Fisher, M. Rizzo, J. K. Caird, & J. D. Lee (Eds.), *Handbook of driving simulation for engineering, medicine, and psychology* (7.1-7.24). CRC Press.

Griggs, T., & Wakabayashi, D. (2018). *How a Self-Driving Uber Killed a Pedestrian in Arizona*. https://www.nytimes.com/interactive/2018/03/20/us/self-driving-uber-pedestrian-killed.html

Habibovic, A., Andersson, J., Nilsson, M., Lundgren, V. M., & Nilsson, J. (2016). Evaluating interactions with non-existing automated vehicles: three Wizard of Oz approaches. In *IEEE Intelligent Vehicles Symposium (IV)* (pp. 32–37). IEEE. https://doi.org/10.1109/IVS.2016.7535360

Habibovic, A., Lundgren, V. M., Andersson, J., Klingegård, M., Lagström, T., Sirkka, A., Fagerlönn, J., Edgren, C., Fredriksson, R., Krupenia, S., Saluäär, D., & Larsson, P. (2018). Communicating Intent of Automated Vehicles to Pedestrians. *Frontiers in Psychology*, *9*(1336). https://doi.org/10.3389/fpsyg.2018.01336

Hagenzieker, M. P., van der Kint, S., Vissers, L., van Schagen, I. N. L. G., Bruin, J. de, van Gent, P., & Commandeur, J. J. F. (2020). Interactions between cyclists and automated vehicles: Results of a photo experiment *Journal of Transportation Safety & Security*, *12*(1), 94–115. https://doi.org/10.1080/19439962.2019.1591556

Hensch, A.-C., Neumann, I., Beggiato, M., Halama, J., & Krems, J. F. (2020a). Effects of a light-based communication approach as an external HMI for Automated Vehicles - a Wizard-of-Oz Study. *Transactions on Transport Sciences*, *10*(2), 18–32. https://doi.org/10.5507/tots.2019.012

Hensch, A.-C., Neumann, I., Beggiato, M., Halama, J., & Krems, J. F. (2020b). How Should Automated Vehicles Communicate? – Effects of a Light-Based Communication Approach in a Wizard-of-Oz Study. In N. Stanton (Ed.), *Advances in Intelligent Systems and Computing, Advances in Human Factors of Transportation* (pp. 79–91). Springer International Publishing. https://doi.org/10.1007/978-3-030-20503-4_8

Hipp, E., Bengler, K., Kressel, U., & Feit, S. (2018). The Research Initiative UR:BAN. In K. Bengler, J. Drüke, S. Hoffmann, D. Manstetten, & A. Neukum (Eds.), *UR:BAN Human Factors in Traffic* (pp. 3–26). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-15418-9_1

Holländer, K., Colley, A., Mai, C., Häkkilä, J., Alt, F., & Pfleging, B. (2019). Investigating the Influence of External Car Displays on Pedestrians' Crossing Behavior in Virtual Reality. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '19* (pp. 1–11). ACM Press. https://doi.org/10.1145/3338286.3340138

Holländer, K., Wintersberger, P., & Butz, A. (2019). Overtrust in External Cues of Automated Vehicles. In Unknown (Ed.), *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '19* (pp. 211–221). ACM Press. https://doi.org/10.1145/3342197.3344528

Hudson, C. R., Deb, S., Carruth, D. W., McGinley, J., & Frey, D. (2019). Pedestrian Perception of Autonomous Vehicles with External Interacting Features. In I. L. Nunes (Ed.), *Advances in Intelligent Systems and Computing. Advances in Human Factors and Systems Interaction* (Vol. 781, pp. 33–39). Springer International Publishing. https://doi.org/10.1007/978-3-319-94334-3_5

Hurwitz, D., Knodler, M., & Dulaski, D. (2005). Speed perception fidelity in a driving simulator environment. In *Conference: Driving Simulator Conference North America,* Orlando, Florida. https://www.researchgate.net/publication/237700597_Speed_Perception_Fidelity_in_a_Driving_Simulator_Environment

Imbsweiler, J., Palyafári, R., Puente León, F., & Deml, B. (2017). Untersuchung des Entscheidungsverhaltens in kooperativen Verkehrssituationen am Beispiel einer Engstelle. *At - Automatisierungstechnik*, *65*(7). https://doi.org/10.1515/auto-2016-0127

Iryo-Asano, M., Hasegawa, Y., & Dias, C. (2018). Applicability of Virtual Reality Systems for Evaluating Pedestrians' Perception and Behavior. *Transportation Research Procedia*, *34*, 67–74. https://doi.org/10.1016/j.trpro.2018.11.015

Jayaraman, S. K., Creech, C., Robert Jr., L. P., Tilbury, D. M., Yang, X. J., Pradhan, A. K., & Tsui, K. M. (2018). Trust in AV. In T. Kanda, S. Ŝabanović, G. Hoffman, & A. Tapus (Eds.), *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18* (pp. 133–134). ACM Press.

Joisten, P., Alexandi, E., Drews, R., Klassen, L., Petersohn, P., Pick, A., Schwindt, S., & Abendroth, B. (2020). Displaying Vehicle Driving Mode – Effects on Pedestrian Behavior and Perceived Safety. In T. Ahram, W. Karwowski, S. Pickl, & R. Taiar (Eds.), *Advances in Intelligent Systems and Computing. Human Systems Engineering and Design II* (Vol. 1026, pp. 250–256). Springer International Publishing. https://doi.org/10.1007/978-3-030-27928-8_38

Kantowitz, B. H. (2011). Using Driving Simulators Outside of North America. In D. L. Fisher, M. Rizzo, J. K. Caird, & J. D. Lee (Eds.), *Handbook of driving simulation for engineering, medicine, and psychology* (3.1-3.14). CRC Press.

Kauffmann, N., Naujoks, F., Winkler, F., & Kunde, W. (2018). Learning the "Language" of Road Users - How Shall a Self-driving Car Convey Its Intention to Cooperate to Other Human Drivers? In I. L. Nunes (Ed.), *Advances in Intelligent Systems and Computing. Advances in Human Factors and Systems Interaction* (Vol. 592, pp. 53–63). Springer International Publishing. https://doi.org/10.1007/978-3-319-60366-7_6

Kauffmann, N., Winkler, F., Naujoks, F., & Vollrath, M. (2018). "What Makes a Cooperative Driver?" Identifying parameters of implicit and explicit forms of communication in a lane change scenario. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 1031–1042. https://doi.org/10.1016/j.trf.2018.07.019

Klebelsberg, D. (1982). *Verkehrspsychologie*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-47507-8

Kühn, M., Stange, V., & Vollrath, M. (2019). Menschliche Reaktion auf hochautomatisierte Fahrzeuge im Mischverkehr auf der Autobahn. In *VDI Tagung Mensch-Maschine-Mobilität – Der (Mit-)Fahrer im 21.Jahrhundert!?* (pp. 169–184). VDI Verlag.

Lagström, T., & Lundgren, V. M. (2015). *AVIP - Autonomous vehicles´ interaction with pedestrians: An investigation of pedestrian-driver communication and development of a* *vehicle external interface* [Master of Science Thesis]. Chalmers University of Technology, Gothenborg.

Lee, Y. M., Madigan, R., Giles, O., Garach-Morcillo, L., Markkula, G., Fox, C., Camara, F., Rothmueller, M., Vendelbo-Larsen, S. A., Rasmussen, P. H., Dietrich, A., Nathanael, D., Portouli, V., Schieben, A., & Merat, N. (2020). Road users rarely use explicit communication when interacting in today's traffic: implications for automated vehicles. *Cognition, Technology & Work*, *5*(9), 145. https://doi.org/10.1007/s10111-020-00635-y

Li, Y., Dikmen, M., Hussein, T. G., Wang, Y., & Burns, C [Catherine] (2018). To Cross or Not to Cross: Urgency-Based External Warning Displays on Autonomous Vehicles to Improve Pedestrian Crossing Safety. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '18* (pp. 188–197). ACM Press. https://doi.org/10.1145/3239060.3239082

Löcken, A., Golling, C., & Riener, A. (2019). How Should Automated Vehicles Interact with Pedestrians? In Unknown (Ed.), *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '19* (pp. 262–274). ACM Press. https://doi.org/10.1145/3342197.3344544

Lüke, S., Fochler, O., Schaller, T., & Regensburger, U. (2015). Stauassistenz und -automation. In H. Winner, S. Hakuli, F. Lotz, & C. Singer (Eds.), *Handbuch Fahrerassistenzsysteme* (F 83, pp. 995–1007). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-05734-3_52

Lundgren, V. M., Habibovic, A., Andersson, J., Lagström, T., Nilsson, M., Sirkka, A., Fagerlönn, J., Fredriksson, R., Edgren, C., Krupenia, S., & Saluäär, D. (2017). Will There Be New Communication Needs When Introducing Automated Vehicles to the Urban Context? In N. A. Stanton, S. Landry, G. Di Bucchianico, & A. Vallicelli (Eds.), *Advances in Intelligent Systems and Computing. Advances in Human Aspects of Transportation* (Vol. 484, pp. 485–497). Springer International Publishing. https://doi.org/10.1007/978-3-319-41682-3_41

Mahadevan, K., Somanath, S., & Sharlin, E. (2018). Communicating Awareness and Intent in Autonomous Vehicle-Pedestrian Interaction. In R. Mandryk, M. Hancock, M. Perry, & A. Cox (Eds.), *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* (pp. 1–12). ACM Press. https://doi.org/10.1145/3173574.3174003

Maruhn, P., Dietrich, A., Prasch, L., & Schneider, S. (2020, March). Analyzing Pedestrian Behavior in Augmented Reality — Proof of Concept. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 313–321). IEEE. https://doi.org/10.1109/VR46266.2020.00051

Matthews, M., Chowdhary, G., & Kieson, E. (2017). *Intent Communication between Autonomous Vehicles and Pedestrians*. http://arxiv.org/pdf/1708.07123v1

Merat, N., Louw, T., Madigan, R., Wilbrink, M., & Schieben, A. (2018). What externally presented information do VRUs require when interacting with fully Automated Road Transport Systems in shared space? *Accident; Analysis and Prevention*, *118*, 244–252. https://doi.org/10.1016/j.aap.2018.03.018

Miguel, M. Á. de, Fuchshuber, D., Hussein, A., & Olaverri-Monreal, C. (2019). Perceived Pedestrian Safety: Public Interaction with Driverless Vehicles. In *2019 IEEE Intelligent Vehicles Symposium (IV)* (pp. 90–95). IEEE.

Moore, D., Currano, R., Strack, G. E., & Sirkin, D. (2019). The Case for Implicit External Human-Machine Interfaces for Autonomous Vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '19* (pp. 295–307). ACM Press. https://doi.org/10.1145/3342197.3345320

Müller, A. I., Weinbeer, V., & Bengler, K. (2019). Using the wizard of Oz paradigm to prototype automated vehicles. In C. P. Janssen, S. F. Donker, L. L. Chuang, & W. Ju (Eds.), *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (pp. 181–186). ACM. https://doi.org/10.1145/3349263.3351526

Neukum, A., Lübbeke, T., Krüger, H.-P., Mayser, C., & Steinle, J. (2008). ACC-Stop&Go: Fahrerverhalten an funktionalen Systemgrenzen. In M. Maurer & C. Stiller (Chairs), *5. Workshop Fahrerassistenzsysteme.* Symposium conducted at the meeting of fmrt, Karlsruhe.

Nimmons, J. (2014). *Effective communication is the most important strategy to achieve your business goals*. http://www.sbnonline.com/article/effective-communication-important-strategy-achieve-business-goals/

Nuñez Velasco, J. P., Farah, H., van Arem, B., & Hagenzieker, M. P. (2019). Studying pedestrians' crossing behavior when interacting with automated vehicles using virtual reality. *Transportation Research Part F: Traffic Psychology and Behaviour*, *66*, 1–14. https://doi.org/10.1016/j.trf.2019.08.015

Othersen, I., Conti-Kufner, A. S., Dietrich, A., Maruhn, P., & Bengler, K. (2019). Designing for Automated Vehicle and Pedestrian Communication: Perspectives on eHMIs from Older and Younger Persons. In D. d. Waard, K. Brookhuis, D. Coelho, S. Fairclough, D. Manzey, A. Naumann, L. Onnasch, S. Röttger, A. Toffetti, & R. Wiczorek (Eds.), *Proceedings of the Ergonomics and Ergonomics Society Europe Chapter 2018 Annual Conference* (pp. 135–148).

Petzoldt, T., Schleinitz, K., & Banse, R. (2017). Laboruntersuchung zur potenziellen Sicherheitswirkung einer vorderen Bremsleuchte in Pkw. *Zeitschrift Für Verkehrssicherheit*, *63*.

Petzoldt, T., Schleinitz, K., & Banse, R. (2018). Potential safety effects of a frontal brake light for motor vehicles. *IET Intelligent Transport Systems*, *12*(6), 449–453. https://doi.org/10.1049/iet-its.2017.0321

Potzy, J., Feinauer, S., Siedersberger, K.-H., & Bengler, K. (2019). Manual Drivers' Evaluation of Automated Merging Behavior in Dense Traffic: Efficiency Matters. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)* (pp. 3454–3460). IEEE. https://doi.org/10.1109/ITSC.2019.8917346

Potzy, J., Feuerbach, M., & Bengler, K. (2019). Communication Strategies for Automated Merging in Dense Traffic. In *2019 IEEE Intelligent Vehicles Symposium (IV)* (pp. 2291–2298). IEEE. https://doi.org/10.1109/IVS.2019.8813835

Przyborski, A., & Wohlrab-Sahr, M. (2014). *Qualitative Sozialforschung*. DE GRUYTER. https://doi.org/10.1524/9783486719550

Rasouli, A., Kotseruba, I., & Tsotsos, J. K. (2018a). Understanding Pedestrian Behavior in Complex Traffic Scenes. *IEEE Transactions on Intelligent Vehicles*, *3*(1), 61–70. https://doi.org/10.1109/TIV.2017.2788193

Rasouli, A., Kotseruba, I., & Tsotsos, J. K. (2018b, November). Towards Social Autonomous Vehicles: Understanding Pedestrian-Driver Interactions. In *21st International*

Conference on Intelligent Transportation Systems (ITSC) (pp. 729–734). IEEE. https://doi.org/10.1109/ITSC.2018.8569324

Rasouli, A., & Tsotsos, J. K. (2018). *Autonomous Vehicles that Interact with Pedestrians: A Survey of Theory and Practice.* http://arxiv.org/pdf/1805.11773v1

Reeves, B., & Nass, C. I. (1997). *The media equation: How people treat computers, televisions, and new media like real people and places* (2nd printing). CSLI Publ.

Rettenmaier, M., Albers, D., & Bengler, K. (2020). After you?! – Use of external human-machine interfaces in road bottleneck scenarios. *Transportation Research Part F: Traffic Psychology and Behaviour*, *70*, 175–190. https://doi.org/10.1016/j.trf.2020.03.004

Rettenmaier, M., & Bengler, K. (2020). Modeling the Interaction with Automated Vehicles in Road Bottleneck Scenarios. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 1615–1619. https://doi.org/10.1177/1071181320641391

Rettenmaier, M., Pietsch, M., Schmidtler, J., & Bengler, K. (2019). Passing through the Bottleneck - The Potential of External Human-Machine Interfaces. In *2019 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1687–1692). IEEE. https://doi.org/10.1109/IVS.2019.8814082

Rettenmaier, M., Schulze, J., & Bengler, K. (2020). How Much Space Is Required? Effect of Distance, Content, and Color on External Human–Machine Interface Size. *Information*, *11*(7), 346. https://doi.org/10.3390/info11070346

Rice, S., Winter, S. R., Doherty, S., & Milner, M. (2017). Advantages and Disadvantages of Using Internet-Based Survey Methods in Aviation-Related Research. *Journal of Aviation Technology and Engineering*, *7*(1). https://doi.org/10.7771/2159-6670.1160

Robert Bosch GmbH. (2020). *Traffic jam assist: Helps to guide the vehicle in traffic jams - relaxed and safely*. https://www.bosch-mobility-solutions.com/en/products-and-services/passenger-cars-and-light-commercial-vehicles/automated-driving/traffic-jam-assist/

Rodríguez, P. (2017). *Safety of Pedestrians and Cyclists when Interacting with Automated Vehicles – A Case Study of the WEpods* [Master of Science Thesis]. Delft University of Technology, Delft.

Rodríguez Palmeiro, A. (2017). *Interaction between pedestrians and Wizard of Oz automated vehicles* [Master Thesis]. Delft University of Technology, Delft. http://repository.tudelft.nl/

Röhner, J., & Schütz, A. (2016). *Psychologie der Kommunikation*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-10024-7

Rothenbücher, D., Li, J., Sirkin, D., Mok, B., & Ju, W. (2015). Ghost Driver: A Field Study Investigating the Interaction Between Pedestrians and Driverless Vehicles. In G. Burnett, J. Gabbard, P. Green, & S. Osswald (Eds.), *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '15* (pp. 44–49). ACM Press. https://doi.org/10.1145/2809730.2809755

Rothenbücher, D., Li, J., Sirkin, D., Mok, B., & Ju, W. (2016, August). Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 795–802). IEEE. https://doi.org/10.1109/ROMAN.2016.7745210

Rouchitsas, A., & Alm, H. (2019). External Human-Machine Interfaces for Autonomous Vehicle-to-Pedestrian Communication: A Review of Empirical Work. *Frontiers in Psychology*, *10, 2757.* https://doi.org/10.3389/fpsyg.2019.02757

SAE (06.2018). *Surface Vehicle Recommended Practice*. (J3016™).

Schmidt, G., Kiss, M., Babbel, E., & Galla, A. (2008). The Wizard on Wheels: Rapid Prototyping and User Testing of Future Driver Assistance Using Wizard of Oz Technique in a Vehicle. In *Proceedings of the FISITA 2008 World Automotive Congress,* Munich.

Schmidt, H., Terwilliger, J., AlAdawy, D., & Fridman, L. (2019). *Hacking Nonverbal Communication Between Pedestrians and Vehicles in Virtual Reality*. https://arxiv.org/pdf/1904.01931.pdf

Schmidt, S [S.], & Färber, B. (2009). Pedestrians at the kerb – Recognising the action intentions of humans. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*(4), 300–310. https://doi.org/10.1016/j.trf.2009.02.003

Schneemann, F., & Gohl, I. (2016). Analyzing driver-pedestrian interaction at crosswalks: A contribution to autonomous driving in urban environments. In *IEEE Intelligent Vehicles Symposium (IV)* (pp. 38–43). IEEE. https://doi.org/10.1109/IVS.2016.7535361

Schneider, S., & Bengler, K. (2020). Virtually the same? Analysing pedestrian behaviour by means of virtual reality. *Transportation Research Part F: Traffic Psychology and Behaviour*, *68*, 231–256. https://doi.org/10.1016/j.trf.2019.11.005

Schneider, S., Maruhn, P., Dang, N.-T., Pala, P., Cavallo, V., & Bengler, K. (2021). Pedestrian Crossing Decisions in Virtual Environments: Behavioral Validity in CAVEs and Head-Mounted Displays. *Human Factors.* Advance online publication. https://doi.org/10.1177/0018720820987446

Schulz von Thun, F. (2018). *Miteinander reden: 1 - Störungen und Klärungen: Allgemeine Psychologie der Kommunikation* (55. Auflage, Originalausgabe). *Rororo: Vol. 17489*. Rowohlt Taschenbuch Verlag.

Sedlmeier, P., & Renkewitz, F. (2011). *Forschungsmethoden und Statistik in der Psychologie* ([Nachdr.]). *PS Psychologie*. Pearson Studium.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shannon, C. E., & Weaver, W. (1998). *The mathematical theory of communication*. Univ. of Illinois Press. https://www.loc.gov/catdir/enhancements/fy1616/98230924-b.html

Shepardson, D. (2020). *Safety driver in fatal Arizona Uber self-driving car crash charged with homicide.* Reuters. https://www.reuters.com/article/uber-selfdriving-idUSKBN26708P

Song, Y. E., Lehsing, C., Fuest, T., & Bengler, K. (2018). External HMIs and Their Effect on the Interaction Between Pedestrians and Automated Vehicles. In W. Karwowski & T. Ahram (Eds.), *Advances in Intelligent Systems and Computing, Intelligent Human Systems Integration* (pp. 13–18). Springer International Publishing. https://doi.org/10.1007/978-3-319-73888-8_3

Vollrath, M. (2015). Statistische Methoden. In H. Bubb, K. Bengler, R. E. Grünen, & M. Vollrath (Eds.), *ATZ / MTZ-Fachbuch. Automobilergonomie* (pp. 663–684). Springer Vieweg.

Wagner, J. (2014). *Diese Gesten machen Sie im Ausland unbeliebt: Achtung, Fettnäpfchen Körpersprache*. https://www.focus.de/reisen/service/tid-33033/andere-laendere-andere-koerpersprache-welche-gesten-im-ausland-unbeliebt-machen_aid_1076395.html

Watzlawick, P. (2016). *Man kann nicht nicht kommunizieren: Das Lesebuch* (2., unveränderte Auflage). Hogrefe.

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. https://doi.org/10.1016/j.jesp.2014.01.005

Weber, F., Chadowitz, R., Schmidt, K., Messerschmidt, J., & Fuest, T. (2019). Crossing the Street Across the Globe: A Study on the Effects of eHMI on Pedestrians in the US, Germany and China. In H. Krömker (Ed.), *Lecture Notes in Computer Science, HCI in Mobility, Transport, and Automotive Systems* (pp. 515–530). Springer International Publishing. https://doi.org/10.1007/978-3-030-22666-4_37

Wiener, N. (2000). *Cybernetics or control and communication in the animal and the machine* (2. ed., 10. print). MIT Press.

Witzlack, C., Beggiato, M., & Krems, J. (2016). Interaktionssequenzen zwischen Fahrzeugen und Fußgängern im Parkplatzszenario als Grundlage für kooperativ interagierende Automatisierung. In VDI (Ed.), *Fahrerassistenz und automatisiertes Fahren. VDI-Berichte 2288* (pp. 323–336). VDI-Verlag.

Wright, K. B. (2005). Researching Internet-Based Populations: Advantages and Disadvantages of Online Survey Research, Online Questionnaire Authoring Software Packages, and Web Survey Services. *Journal of Computer-Mediated Communication*, *10*(3). https://doi.org/10.1111/j.1083-6101.2005.tb00259.x

Yang, S. (2017). *Driver behavior impact on pedestrians' crossing experience in the conditionally autonomous driving context* [Degree Project in Computer Science and Communication]. KTH Royal Institute of Technology, Stockholm, Sweden.

Zhang, J., Vinkhuyzen, E., & Cefkin, M. (2018). Evaluation of an Autonomous Vehicle External Communication System Concept: A Survey Study. In N. A. Stanton (Ed.), *Advances in Intelligent Systems and Computing. Advances in Human Aspects of Transportation* (Vol. 597, pp. 650–661). Springer International Publishing. https://doi.org/10.1007/978-3-319-60441-1_63