Context-Based Prototyping of Human-Machine Interfaces for Autonomous Vehicles

A dissertation submitted towards the degree
Doctor of Engineering (Dr.-Ing.)
of the Faculty of Mathematics and Computer Science
of Saarland University

Lukas A. Flohr
Saarbrücken, 2023

Saarland University
Faculty of Mathematics and Computer Science
Department of Computer Science

Context-Based Prototyping of Human-Machine Interfaces for Autonomous Vehicles

A dissertation submitted towards the degree
Doctor of Engineering (Dr.-Ing.)
of the Faculty of Mathematics and Computer Science
of Saarland University

Dissertation zur Erlangung des Grades
des Doktors der Ingenieurwissenschaften (Dr.-Ing.)
der Fakultät für Mathematik und Informatik
der Universität des Saarlandes

Lukas Amadeus Flohr, M. Sc.
Saarbrücken, 2023

DEAN	Prof. Dr. Jürgen Steimle
EXAMINATION BOARD	
Reporter	Prof. Dr. Antonio Krüger Prof. Dr. Dieter Wallach Prof. Dr. Andreas Butz
Chair	Prof. Dr. Jörg Hoffmann
Assistant	Dr. Donald Degraen
DATE OF THE COLLOQUIUM	13 October 2023

NOTES ON COLLABORATION, STYLE, AND PREVIOUS PUBLICATIONS

This doctoral research was conducted in cooperation with the Ubiquitous Media Technology Lab (UMTL) at Saarland Informatics Campus and Ergosign GmbH. Large parts of the work presented in this dissertation were done in collaboration with colleagues, other researchers, and students. Therefore, the scientific plural "we" is used predominantly in this document. For statements only referring to my personal experiences and perspectives, I will use the first-person singular.

Academic publications achieved during the doctoral research provide the basis for this dissertation. A complete publication list and corresponding author contributions can be found on pp. ix—xiii.

KEYWORDS

Context-based interface prototyping; autonomous vehicles; human-machine interfaces; human-centered design; prototyping methods; simulation; wizard-of-oz; physical context; social context; augmented reality; acceptance; user experience.

Acknowledgements

It has been a ride. One that would not have been possible without the support of a lot of people. At this point in my journey, I would like to mention and thank some of those.

First and foremost, I would like to thank Prof. Dr. Antonio Krüger and Prof. Dr. Dieter Wallach for privileging me to pursue this doctoral research meanwhile enabling me to travel both the academic and the business world. Antonio, thank you for your supervision, support, and invaluable ideas and for accepting me as an external doctoral researcher at UMTL. Dieter, thank you for being the mentor I needed during my journey. You are one of the most impressive and inspiring persons I have ever had the pleasure of working with. Thank you for all your efforts, the inspirational conversations, and the occasional joke about what happened last night.

Furthermore, I would like to express my gratitude to Prof. Dr. Andreas Butz for providing the third review of this dissertation. I feel honored and hope you enjoyed reading it. I also want to thank Prof. Dr. Jörg Hoffmann for chairing the examination board and Dr. Donald Degraen for serving as the board's scientific assistant, as well as Prof. Dr. Jürgen Steimle for being my academic mentor at Saarland University.

I have had the honor and pleasure of working with amazing people during my doctoral research. Many of them also became close friends of mine. A big thank you goes to my colleagues at Ergosign. I would like to thank Sebastian C. Scholz for supporting me particularly in the starting phase of my doctoral research with his enthusiasm, trust, and ideas. And yes, it takes ~~talent~~ tenacity. Similarly, I would like to thank my team managers Stefan Kiefer and Stefan Schulz for their trust, feedback, and support. Also, a huge thank you to Juliane Zirke, my office roommate during my journey, who always had an open ear and valuable perspectives for me. I am also grateful to have been working with some talented colleagues and students. Especially, I want to mention Julian Schneider, Dominik Janetzko, Sofie Kalinke, and Joseph Valiyaveettil, who significantly contributed to this work with their dedication and extraordinary commitment. Furthermore, I want to thank Annika Kaltenhauser and Amelie Nolte for sharing my fate as a doctoral researcher at Ergosign and for all the invaluable and motivating discussions across our domains. I would also like to thank Dr. Verena Rheinstädter for her support and valuable feedback. Thank you, Michael Alexander, for your excellent ideas and hands-on support in creating our Wizard-of-Oz vehicle setup. Another big thank you goes to the UMTL folks for welcoming me in their research family and all the inspiring Friday Talks. Particularly, I want to thank Dr. Pascal Lessel, Dr. André Zenner, Iris Lambrecht, and Gundula Kleiner for their support and for answering my many organizational questions. Also, thank you, Anke Hirsch, Nina Knieriemen, and Frederik Wiehr, for the delight of co-conducting our human-centered design seminar. I want to thank Martina Schuß for a fantastic collaboration across the 'Tellerrand' of companies and universities. Thank you for the mutual trust, hard work, and motivation. Here, I also want to thank Prof. Dr. Andreas Riener, who connected me with Martina after our conversations at the GermanUPA's Automotive Systems working group and his proceeding support of our joint research.

From time to time, it can be rough riding the doctoral research train. I count myself happy that I have amazing friends and a caring family that accompanied me through this journey's ups and downs. A particular thank you goes to Dominik Janetzko and Deike Albers, who have shared a similar journey with me since our student days. Thanks for all the professional discussions on study designs, ethics, and statistics (Dominik is basically our personal guru in this matter), and, of course, the pep talks, and the pleasureful talks after one or the other beer. An enormous thank you also goes to the people who gifted me with their valuable time and proofread parts of this thesis, particularly Alexander, Deike, Dominik, Esther, Fabian, Florian, Stefan, and Verena.

Aus tiefstem Herzen möchte ich den liebsten Menschen in meinem Leben danken. Dank meiner Eltern, Franziska und Berthold, und ihrer fördernden und toleranten Erziehung, war es mir möglich eine akademische Laufbahn zu verfolgen. Ihr gabt mir stets den Rückenwind meine eigene Welt zu entdecken, aber zeitgleich das Gefühl, dass ihr jederzeit für mich da seid. Danke für Alles! Ein großer Dank gebührt zudem meinen Brüdern, Fabian und Benjamin, sowie ihren wunderbaren Familien. Rückblickend wäre ich ohne Fabian vermutlich nie auf die Idee gekommen eine Promotion anzustreben. Als kleiner Bruder orientiert man sich wohl doch stark an den Großen. Danke für eure Unterstützung, eure Vorbildfunktion und die vielen Gespräche zu den großen und kleinen Dingen des Lebens. Schließlich gebührt der größte Dank meiner wundervollen Anne-Sophie. Du bist und bleibst meine Lieblings-Reisebegleiterin. Danke für deine Liebe, dein Verständnis, deine Unterstützung von Tag zu Tag, dein herausforderndes Wesen und deine ermutigenden Worte. Du bist die Beste!

Finally, I want to thank Ergosign GmbH, the UMTL, and the Saarbrücken Graduate School of Computer Science for providing me with this opportunity. Parts of my research have received funding from the German Federal Ministry of Transport and Digital Infrastructure (BMVI) under grant number 16AVF2134G (project APEROL) and from the German Federal Ministry of Economic Affairs and Climate Action (BMWK) under grant number 19A22006K (project STADT:up) for which I am very grateful.

Abstract

Autonomous vehicles (AVs; SAE levels 4 and 5) face substantial challenges regarding acceptance, human factors, and user experience. Human-machine interfaces (HMIs) offer the potential to account for those and facilitate AV adoption. Since AVs′ capabilities and availability are still limited, suitable prototyping methods are required to create, evaluate, and optimize novel HMI concepts from early development phases. In all human-centered design activities, physical and social contexts are vital. This thesis argues for applying context-based interface prototyping of human-AV interactions to account for their interrelation with contextual factors. We adopt a ′research in and through design′ approach and explore the two intertwined areas: design and prototyping. Regarding the latter, we concentrate on straightforward methods. We demonstrate an immersive video-based approach for lab simulation of AVs and a wizard-of-oz-based method for on-road AV simulation and prototyping of HMIs providing real-time information. We apply these methods in empirical studies to assess their suitability and explore HMI concepts created to counter the aforementioned challenges. Thereby, we investigate the potential of (AR-based) object detection visualization and concepts for mobile and in-vehicle interaction with (shared) AVs. Based on the findings, we provide design and prototyping recommendations that will aid researchers and practitioners in creating suitable human-AV interactions.

Zusammenfassung

Autonome Fahrzeuge (AVs; SAE Level 4 und 5) stehen vor großen Herausforderungen in Bezug auf Akzeptanz, Human Factors und User Experience. Mensch-Maschine-Schnittstellen (HMIs) haben das Potenzial, diesen entgegenzutreten und die Einführung fahrerloser Fahrzeuge zu erleichtern. Da Fähigkeiten und Verfügbarkeiten von AVs noch begrenzt sind, sind geeignete Prototyping-Methoden erforderlich, um neue HMI-Konzepte bereits in frühen Entwicklungsphasen zu erstellen, zu evaluieren und zu optimieren. Bei deren menschzentrierter Gestaltung sind physische und soziale Kontexte entscheidend. Zur frühzeitigen Berücksichtigung der Wechselbeziehung von HMIs mit kontextuellen Einflussfaktoren wird in dieser Thesis für die Anwendung von kontextbasiertem Interface-Prototyping von Mensch-AV-Interaktionen plädiert. Wir verfolgen einen ′research in and through design′-Ansatz und untersuchen die beiden miteinander verknüpften Bereiche: Design und Prototyping. Bei Letzterem konzentrieren wir uns auf einfache, aber effektive Methoden. Wir demonstrieren einen immersiven video-basierten Ansatz zur Laborsimulation von (geteilten) AVs und eine ′Wizard of Oz′-basierte Methode für die Simulation autonomer Fahrten auf realen Straßen und das Prototyping von HMIs mit Echtzeitinformationen. Wir wenden diese Methoden in empirischen Studien an, um ihre Eignung zu bewerten und um HMI-Konzepte zu untersuchen, die zur Bewältigung der genannten Herausforderungen entwickelt wurden. Dabei untersuchen wir das Potenzial von (AR-basierter) Visualisierung von Objekterkennungen und Konzepten für die mobile und fahrzeuginterne Interaktion mit (geteilten) AVs. Basierend auf den Ergebnissen geben wir Design- und Prototyping-Empfehlungen, die Forschern und Praktikern helfen sollen, geeignete Mensch-AV-Interaktionen zu entwickeln.

List of Publications

This dissertation is based on and uses parts of publications achieved during the doctoral research. The author's publications and related artifacts, such as talks, workshops, and advised student theses, are listed chronologically below. They are supplemented by descriptions of the individual contributions of authors (abbreviated with their initials) and supporters. Works used in this dissertation are marked with references to the respective chapters in which they appear.

Full Conference Papers

[85] **Lukas A. Flohr**, Dominik Janetzko, Dieter P. Wallach, Sebastian C. Scholz, and Antonio Krüger. 2020. Context-Based Interface Prototyping and Evaluation for (Shared) Autonomous Vehicles Using a Lightweight Immersive Video-Based Simulator. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1379–1390. `https://doi.org/10.1145/3357236.3395468`
→ APPEARS IN CHAPTERS 1, 2, **3**, AND 7

Contributions: conceptualization, L.A.F. and S.C.S..; methodology, L.A.F. and D.J.; data analysis, L.A.F. and D.J.; writing—original draft preparation, L.A.F.; writing—review and editing, D.J., D.P.W., S.C.S., A.K., and L.F.; visualization, L.A.F.; supervision, S.C.S, D.P.W., and A.K.; project administration, L.A.F. and D.P.W.; funding acquisition, D.P.W.; Our colleagues Julian Schneider and Florian Schneider supported the original simulator's design and construction. Julian also co-conducted the reported expert study. Verena Rheinstädter and Annika Kaltenhauser supported us with valuable discussions and feedback on the original paper draft.

[260] Dieter P. Wallach, **Lukas A. Flohr**, and Annika Kaltenhauser. 2020. Beyond the Buzzwords: On the Perspective of AI in UX and Vice Versa. In *Proceedings of the 1st International Conference on Artificial Intelligence in HCI, Held as Part of the 22nd International Conference on Human-Computer Interaction* (Copenhagen, Denmark) *(HCII '20)*. Springer International Publishing, Cham, Switzerland, 146–166. `https://doi.org/10.1007/978-3-030-50334-5_10`
→ APPEARS IN CHAPTERS 1, 2, and 7

Contributions: conceptualization, D.P.W., L.A.F., and A.K.; methodology, D.P.W., L.A.F., and A.K.; writing—original draft preparation, L.A.F., D.P.W., and A.K.; writing—review and editing, D.P.W., A.K., and L.A.F.; visualization, L.A.F., A.K., and D.P.W.; supervision, D.P.W.; project administration, D.P.W.; funding acquisition, D.P.W.

[86] **Lukas A. Flohr**, Sofie Kalinke, Antonio Krüger, and Dieter P. Wallach. 2021. Chat or Tap? – Comparing Chatbots with 'Classic' Graphical User Interfaces for Mobile Interaction with Autonomous Mobility-on-Demand Systems. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. `https://doi.org/10.1145/3447526.3472036`
→ APPEARS IN CHAPTERS 1, 2, **5**, AND 7

Contributions: conceptualization, L.A.F. and S.K..; methodology, L.A.F. and S.K.; data analysis, S.K. and L.A.F.; writing—original draft preparation, L.A.F.; writing—review and editing, S.K., D.P.W., A.K., and L.A.F.; visualization, L.A.F. and S.K.; supervision, D.P.W. and A.K.; project administration, L.A.F.; funding acquisition, D.P.W.

[88] **Lukas A. Flohr**, Joseph S. Valiyaveettil, Antonio Krüger, and Dieter P. Wallach. 2023. Prototyping Autonomous Vehicle Windshields with AR and Real-Time Object Detection Visualization: An On-Road Wizard-of-Oz Study. *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. `https://doi.org/10.1145/3563657.3596051`
→ APPEARS IN CHAPTERS 1, 2, **4**, AND 7

Contributions: conceptualization, L.A.F.; methodology, L.A.F. and J.S.V.; software development: J.S.V. with support from Björn Busch-Geertsema; data analysis: L.A.F. with support from J.S.V., Ria Glink, Hanna Schumm, and Barbara Gleißl; writing—original draft preparation, L.A.F.; writing—review and editing, J.S.V., D.P.W., A.K., and L.A.F.; visualization, L.A.F.; supervision, D.P.W., and A.K.; project administration, L.A.F.; funding acquisition, D.P.W. and L.A.F.; Our colleague Michael Alexander supported us in the mechanical study setup; Sofie Kalinke supported us in the study conduct.

Journal Papers

[203] Natalie T. Richardson, **Lukas Flohr**, and Britta Michel. 2018. Takeover Requests in Highly Automated Truck Driving: How Do the Amount and Type of Additional Information Influence the Driver–Automation Interaction? *Multimodal Technologies and Interaction* 2, 4 (2018), 68. `https://doi.org/10.3390/mti2040068`

Contributions: conceptualization, N.T.R. and L.F.; methodology, N.T.R. and L.F.; data curation, N.T.R.; formal analysis, L.F. and N.T.R.; investigation, N.T.R. and L.F.; validation, N.T.R. and L.F.; writing—original draft, N.T.R.; writing—review and editing, B.M.; visualization, L.F.; supervision, N.T.R.; project administration, N.T.R. and L.F.; funding acquisition, N.T.R.

[89] **Lukas A. Flohr** and Dieter P. Wallach. 2023. The Value of Context-Based Interface Prototyping for the Autonomous Vehicle Domain: A Method Overview. *Multimodal Technologies and Interaction* 7, 4 (2023), 1–17. `https://doi.org/10.3390/mti7010004`

→ APPEARS IN CHAPTERS 1, 2, AND 7

Contributions: conceptualization, L.A.F. and D.P.W.; methodology, L.A.F.; writing—original draft preparation, L.A.F.; writing—review and editing, D.P.W. and L.A.F.; visualization, L.A.F.; supervision, D.P.W.; project administration, L.A.F.; funding acquisition, D.P.W.

Under Review / Preprint

[87] **Lukas A. Flohr***, Martina Schuß*, Dieter P. Wallach, Antonio Krüger, and Andreas Riener. 2023. Designing for Passengers' Information Needs on Fellow Travelers: A Comparison of Day and Night Rides in Shared Automated Vehicles. *arXiv* 2308.02616 (aug 2023), 22 pages. `https://doi.org/10.48550/arXiv.2308.02616`

→ APPEARS IN CHAPTERS 1, 2, **6**, AND 7

Contributions: conceptualization, L.A.F.* and M.S.*; methodology, L.A.F.* and M.S.*; data analysis, L.A.F.* and M.S.* with support from Tatjana Röhr; writing—original draft preparation, L.A.F.* and M.S.*; writing—review and editing, A.R., D.P.W., A.K., L.A.F.*, and M.S.*; visualization, L.A.F.* and M.S.*; supervision, A.R., D.P.W., and A.K.; project administration, L.A.F.* and M.S.*; funding acquisition, A.R. and D.P.W.; Claus Pfeilschifter provided us with technical support before and during the study. Tatjana Röhr also supported us in conducting the study.

**Note: both authors contributed equally to this paper.*

Talks, Presentations, and Workshops

- Speaker at the *Mucbook Creative Night 2020* held as part of the *Munich Creative Business Week 2020 (MCBW '20)*. Kollaborative und menschzentrierte Entwicklung künftiger Mobilitätskonzepte. Munich, Germany, March 2020.
- Paper presentation at the *ACM Designing Interactive Systems Conference 2020 (DIS '20)*. Context-Based Interface Prototyping and Evaluation for (Shared) Autonomous Vehicles [. . .]. Online, July 2020.
- Co-speaker with Annika Kaltenhauser and Dieter P. Wallach at the *Usability in Germany Tagung 2020 (UIG '20)*. Mehr als nur Buzzwords: Zum Verhältnis von KI und UX. Online, November 2020.
- Speaker at the *GermanUPA AK Automotive Systems* workshop on "Usability Drives Mobility" at *ACM Mensch und Computer 2020 (MuC '20)*, online, September 2020.
- Co-speaker with Daniel Lauer at the nationwide *Digitaltag 2021*. Autonomes Fahren & Ich – Ein Blick auf die fahrerlose Zukunft. Online, June 2021.
- Co-speaker with Daniel Kerpen and Dieter P. Wallach at the *Usability in Germany (UIG) Tagung 2021 (UIG '21)*. Prototyping im UX Design: Varianten, Funktionen – und Missverständnisse. Online, September 2021.
- Paper presentation at the *23rd International Conference on Mobile Human-Computer Interaction (MobileHCI '21)*. Chat or Tap? [. . .]. Online, September 2021.
- Co-organizer of the *GermanUPA Arbeitskreis Automotive Systems* workshop on "Future Mobility: Interacting with Autonomous Mobility-on-Demand" at the *ACM Mensch und Computer Conference 2021 (MuC '21)*. Online, September 2021.
- Speaker at the *Doctoral Consortium 2021 of the DGPs Fachgruppe Verkehrspsychologie*. User Interfaces for Autonomous Mobility-on-Demand. Braunschweig, Germany, September 2021.
- Speaker at the *Doctoral Consortium 2022 of the DGPs Fachgruppe Verkehrspsychologie*. Context-Based Prototyping of Autonomous Vehicle HMIs. Dresden, Germany, September 2022.
- Paper presentation at the *ACM Designing Interactive Systems Conference 2023 (DIS '23)*. Prototyping Autonomous Vehicle Windshields with AR and Real-Time Object Detection Visualization [. . .]. Pittsburgh, PA, USA, July 2023.

Student Theses

During my doctoral research, I have had the honor and pleasure of mentoring and advising some talented students on their final theses. At the time of submission, two bachelor's and three master's theses were completed. Below, those are listed chronologically, along with references to corresponding chapters and publications to which some have contributed.

- Julian Schneider. 2019. Fahrgastinformationssyteme für autonome Shuttles in Mobility on Demand Services. *Bachelor's Thesis.* Trier University of Applied Sciences. → CONTRIBUTES TO CHAPTER 3 AND PUBLICATION [85]
- Sofie Kalinke. 2020. Change of Plans – Comparison of Graphical and Conversational User Interfaces in Autonomous Mobility-on-Demand Systems in Case of Changing User Requirements. *Master's Thesis.* Technical University of Munich. → CONTRIBUTES TO CHAPTER 5 AND PUBLICATION [86]
- Joseph Sebastian Valiyaveettil. 2022. User Centric Development for AI Applications: Object Detection and Multi-Object Tracking in User Experience Research of Autonomous Vehicles. *Master's Thesis.* Rosenheim Technical University of Applied Sciences. → CONTRIBUTES TO CHAPTER 4 AND PUBLICATION [88]
- Alexander Altmaier. 2023. Analyse, Konzeption und prototypische Umsetzung eines SAE Level 4 Shuttle Fahrgastinfodisplays. *Bachelor's Thesis.* University of Applied Sciences Kaiserslautern.
- Tiana Meiers. 2023. Investigating the Influence of Provided Information on Users' Willingness to Share Rides in Autonomous Mobility-on-Demand Systems. *Master's Thesis.* Saarland University.

Contents

1

Introduction

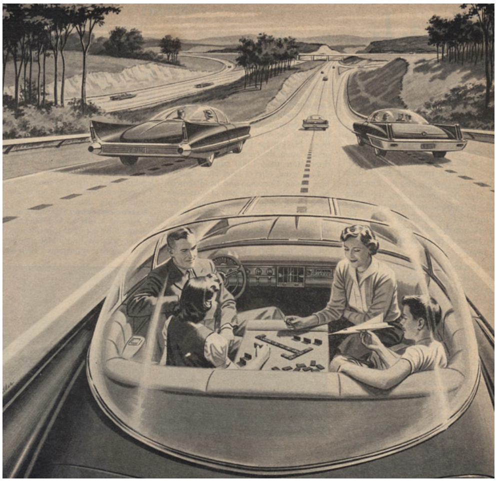


Figure 1.1 – Driverless and accident-free travel. The vision of autonomous travel has been around for almost a century. Image taken from Kröger [144, p. 51]; original source: Americas Independent Electric Light and Power Companies, Advertising, In: LIFE Magazine Vol. 40, N° 5, 30. January 1956, p. 8.

Humanity's dream of autonomous mobility dates back almost a century [144] (Figure 1.1). However, its fulfillment always continued to stay about 20 years away [144]. With recent developments, it seems a tipping point in automotive history is reached. Autonomous vehicles (AVs) are increasingly deployed on test tracks and even tested on public roads all over the world – e.g., in the United States [112, 206, 264], in the United Kingdom [104], in China [137], or in Germany [181]. Although most test operations are still conducted with limitations – e.g., in terms of low-speed limits, restricted areas like airports or university campuses, limited operation times, or special safety regulations like the continuous oversight of a safety/backup driver – it seems that the time gap to the 'always

THIS CHAPTER IS BASED ON THE FOLLOWING PREVIOUS PUBLICATIONS. PARTS OF THIS CHAPTER WERE PUBLISHED PREVIOUSLY AS PART OF THEM.

[85] **Lukas A. Flohr**, Dominik Janetzko, Dieter P. Wallach, Sebastian C. Scholz, and Antonio Krüger. 2020. Context-Based Interface Prototyping and Evaluation for (Shared) Autonomous Vehicles Using a Lightweight Immersive Video-Based Simulator. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1379–1390. `https://doi.org/10.1145/3357236.3395468`

[86] **Lukas A. Flohr**, Sofie Kalinke, Antonio Krüger, and Dieter P. Wallach. 2021. Chat or Tap? – Comparing Chatbots with 'Classic' Graphical User Interfaces for Mobile Interaction with Autonomous Mobility-on-Demand Systems. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. `https://doi.org/10.1145/3447526.3472036`

[87] **Lukas A. Flohr***, Martina Schuß*, Dieter P. Wallach, Antonio Krüger, and Andreas Riener. 2023. Designing for Passengers' Information Needs on Fellow Travelers: A Comparison of Day and Night Rides in Shared Automated Vehicles. *arXiv* 2308.02616 (aug 2023), 22 pages. `https://doi.org/10.48550/arXiv.2308.02616`.
**Note: both authors contributed equally to this paper.*

[88] **Lukas A. Flohr**, Joseph S. Valiyaveettil, Antonio Krüger, and Dieter P. Wallach. 2023. Prototyping Autonomous Vehicle Windshields with AR and Real-Time Object Detection Visualization: An On-Road Wizard-of-Oz Study. *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. `https://doi.org/10.1145/3563657.3596051`

[89] **Lukas A. Flohr** and Dieter P. Wallach. 2023. The Value of Context-Based Interface Prototyping for the Autonomous Vehicle Domain: A Method Overview. *Multimodal Technologies and Interaction 7*, 4 (2023), 1–17. `https://doi.org/10.3390/mti7010004`

[260] Dieter P. Wallach, **Lukas A. Flohr**, and Annika Kaltenhauser. 2020. Beyond the Buzzwords: On the Perspective of AI in UX and Vice Versa. In *Proceedings of the 1st International Conference on Artificial Intelligence in HCI, Held as Part of the 22nd International Conference on Human-Computer Interaction* (Copenhagen, Denmark) *(HCII '20)*. Springer International Publishing, 146–166. `https://doi.org/10.1007/978-3-030-50334-5_10`

20-year away utopia' will diminish [144]. While 2015's expectations [205] estimating driverless vehicles to become a major form of transport until 2025 still seem too optimistic at the time of writing this thesis, emerging commercial test operations with driverless vehicles – e.g., in Phoenix and San Francisco [206, 266] – indicate that the automotive industry might be able to fulfill that promise within the next decade.

Besides the impressive developments of the last years, current test operations also reveal pitfalls and problems that need to be solved for the technology to succeed. For instance, recent tests of AVs in the United States were reported to block streets, lock passengers inside the vehicles, and disregard traffic rules and authorities' instructions [280]. Technical problems aside, AVs face significant challenges in the field of human-computer interaction (HCI). These include issues and concerns related to user acceptance, trust, user experience (UX), safety, and privacy [24, 30, 132, 193, 195, 228]. Future mobility systems offering driverless rides need to compensate prospective passengers' concerns due to the absence of a human driver and overcome those acceptance challenges for AVs to become a major form of transportation.

Counteracting the challenges in the field of human-AV interaction served as one of the key motivations for the works conducted within this doctoral research. In particular, we emphasize the potential of context-based prototyping and design of suitable human-machine interfaces (HMIs) to support the achievement of this goal from the early development phases. In this chapter, we provide a detailed introduction to 1) the era of vehicle automation, 2) the relationship between humans and AVs, and 3) the nature of prototyping, followed by an intro to 4) the thesis' motivation and the research questions pursued, and provide 5) an outline of the thesis structure and applied methodology.

1.1 The Era of Vehicle Automation

Generally, machines are increasingly taking over tasks that humans have previously performed – a process commonly described as automation [189]. Wickens et al. [269] identify four general reasons to automate systems and processes: 1) tasks are dangerous or not possible to conduct for humans, 2) tasks are (too) challenging for the (un-aided) conduct by human operators, 3) tasks conducted by human operators benefit from the support of automation (and consequently optimize the performance of the overall system [202]), and 4) simply because it is inexpensive or possible from a technological point of view. Regarding the automation of road vehicles, all of these reasons appear to be valid.

With the increasing amount of software implemented in today's vehicles and with regard to more and more technology companies (e.g., Alphabet's Waymo [197]) entering the automotive sector, vehicles become 'computers on wheels' [15, 74]. Beiker et al. [15] assigned in 2016 four supporting trends to this development: 1) electrification, 2) autonomous driving, 3) diverse mobility, and 4) connectivity. As a result, they concluded in 2016 that a new ecosystem of the convergence of automotive and tech [15] would evolve. About seven years later, the rising availability of electric and hybrid vehicles, advanced driver assistance systems, and shared and "new" mobility concepts seem to confirm their expectations.

Automation can increase safety, efficiency, and comfort in road traffic [18]. Automated vehicles are expected to reduce traffic jams [238] and to lower air pollution [251]. Vehicle automation has a large potential to contribute to the sustainable development goals of the United Nations (see [254]) – in particular to goal 11 "sustainable cities and communities". According to the World Health Organization, between 20 and 50 million people are injured non-fatally in road traffic, and over 1.3 million people are injured fatally yearly [273, 274]. The European Commission has set itself the goal to reduce road deaths on European streets to zero by 2050 [78]. Considering the fact that the large majority of road traffic accidents can be accounted to human error [217, 241], there is an enormous potential in vehicle automation.

With expanding automation, the role of humans in vehicles is radically changing. Humans transform from active operators and decision-makers to passive passengers [272] who might use their gained free time for non-driving related activities (e.g., communication, productivity, or relaxation [73]). The Society of Automotive Engineers (SAE) and the International Organization for Standardization (ISO) describe six levels – from 0 (no driving automation) to 5 (full driving automation) – to categorize the degree of vehicle automation [210] (Figure 2.1). Initial (advanced driver-assistance) systems with SAE level 3 are already available for privately owned vehicles. They can fully take over the primary driving tasks in specific situations (e.g., during congestion on highways [166]).

However, driverless AVs – i.e., vehicles with SAE levels 4 (high driving automation) and 5 (full driving automation) [210] – are the more significant lever to the technology's potential. Pairing AVs with on-demand ride-sharing enables the creation of autonomous mobility-on-demand (AMoD) systems [191]. In 2022, the first shared AVs started to become a commercial mode of transportation, e.g., [206, 265]. Simulations suggest that shared AVs can meet personal transportation needs in metropolitan cities with only one-third of the current operating passenger vehicles [238]. This expectation illustrates the technology's tremendous economic, societal, and ecological potential. Furthermore, since vehicles can be shared among different people in a staggered way, the amount of traffic resulting from the search for parking lots will be reduced. So does the need for parking lots, which eventually frees up valuable urban space previously occupied by parking traffic.

Comparable to taking a ride in a (shared) taxi, (shared) AV and AMoD journeys provide temporal and spatial flexibility. This means, they neither require fixed timetables nor fixed pick-up or drop-off locations. However, given that there will not be a driver or another accompanying human assistant available within the AV (e.g., to answer traveler queries), AMoD varies decisively from current mobility-on-demand services. The new situation of riding in a driverless vehicle might feel awkward to passengers exposed to an autonomous system's decisions and actions. Thus, digital HMIs capable of filling the resulting service and information gap are needed. Such HMIs require to compensate for the absence of a human driver and to gain users' trust and public acceptance — which is, besides technological hurdles, a main challenge of AVs [132].

1.2 Humans and Autonomous Vehicles

In AVs, the role of humans entirely shifts to passive passengers without control over the primary driving tasks. Consequently, passengers need to accept this unfamiliar and potentially awkward situation of being exposed to an artificial intelligence (AI) powered system's actions and decisions. Previous work has identified trust as a critical challenge for users' acceptance of driverless AVs rides [132]. Further identified issues are related to "concerns about safety, security, usability, accessibility, and comfort" [195]. To successfully adopt the technology, those challenges must be overcome [132].

Since vehicle automation is becoming more complex and interconnected, Lacher et al. [149] conclude that a clear understanding of people, systems, and their interaction in a particular environment is required. Developing AVs in terms of human-centered artificial intelligence (HCAI) [204, 234] can provide an adequate framework to achieve this goal. Besides comprehending humans, their abilities, and needs, a critical aspect of human-centered AI is "to help humans understand AI systems" [204]. Transparent communication can serve as the basis for this understanding and, consequently, increase people's confidence and willingness to use these systems [204].

When a human driver is no longer required, two things need to be considered. First, all occupants are passengers who do not have to pay attention to the driving task. Thus, they can perform non-driving related activities while traveling. Similar to train or bus travels, passengers might use the attained flexibility for communication, productivity or relaxation [73]. Second, human-machine interfaces (HMIs) must provide the passengers with all the information they need along their journey. Such HMIs may compensate (partly) for the absence of a human driver and can become a key source of passengers' trust in the system. This is particularly important for shared autonomous vehicles (SAVs).

SAVs could solve common challenges of today's public transport systems, e.g., regarding congestion, accessibility or first and last mile problems [44, 115, 191, 238]. They promise to provide low-cost traveling and, as vehicle stops can be spatially and temporally flexible, SAVs can potentially substitute the demand for personal cars [153, 191]. In addition to technical hurdles that need to be overcome, a lack of public trust is considered as the central barrier for adaption of SAVs [132]. To defy these hurdles, human factors and user requirements need to be considered from early development phases on [31]. Suitable analysis, design and evaluation methods, as well as appropriate prototyping approaches, are required to inform and enable researchers as well as system designers, developers, and stakeholder such as city municipals or transportation service providers.

Most current HCI research in the field of AMoD focuses on general acceptance aspects of AVs (e.g., [24, 181, 41]). Only a few studies put a particular focus on the human-AV interaction (e.g., [69, 135]). However, for the broad market introduction of AVs and AMoD systems, the HCI community needs to be able to provide future practitioners, manufacturers and service providers with guidelines and recommendations for human-centered design (HCD). Suitable prototyping methods are required to develop systems that defy the hurdles mentioned above. The following section provides the background motivation for creating and using prototypes and elaborates on how they support tackling complex challenges by incorporating context in the design process.

1.3 On the Nature of Prototyping

In a nutshell, the term prototyping describes the creation of (pre-final) representations of (or parts of) a product, system, or service [124]. Proverbs such as "if a picture is worth a thousand words, then a prototype is worth 10,000" [262, p. 5], point out that prototypes not only show and tell, but make ideas, designs, and artifacts tangible [262]. Throughout the process of prototyping, the principle of "learning by doing" is essential [243]. We agree with the broad view of Thaler [247], who concluded that a prototype can basically be "anything that will move the process forward".

In product design, a prototype usually refers to "a pre-production representation of some aspect of a concept or final design" [40, p. 1]. Service prototypes are used to simulate (already existing, (not yet) available, or new) service experiences and enable the consideration of relevant aspects of the real world environment [243]. In the field of UX design, a prototype is referred to as something that "captures the intent of a design and simulates multiple states of that design" [262, p. xii]. Prototypes have a large impact on the success of design and development projects [40]. By making experiences tangible, prototyping reduces misinterpretations, can save time, effort, and money, and reduces the amount of waste created in the process [262]. As Camburn et al. [40] elaborate, prototypes can serve various objectives in product development, including, but not limited to, refinement, exploration, communication, learning, and – in terms of economic perspectives – cost or time reduction. Prototypes empower designers, researchers, users, and other stakeholders (1) to understand context and users' experiences, (2) to explore and evaluate new approaches, and (3) to communicate ideas [35, 85].

Generally, there are two primary use cases of prototypes often distinguished: supporting the generation or exploration of ideas for designing interfaces and evaluating the quality of ideas, concepts, and solutions, especially in early development stages [14]. Often, prototypes are just considered for the latter. However, as Lim et al. [154] pointed out, prototypes can be a tool for "traversing a design space" to gain knowledge about the envisioned product or system and also serve as "manifestations of design ideas" [154]. Through the consideration of both use cases (generation and evaluation), prototyping becomes an essential component in the design process that supports informed decision making.

Depending on their realization, most prototypes are typically limited in some ways, e.g., in their implemented functionality or fidelity (i.e., level of detail [58]). Despite their limitations, prototypes can be used for several major activities, such as design and evaluation, but also for analysis [124] and are thus particularly valuable in the human-centered design of products, services, and interactive systems. Prototypes are (always) used in a particular context [124]. In HCI, context is usually regarded as *"[. . .] any information that can be used to characterize the situation of an entity [. . . which can refer to] a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves"* [65, pp. 3–4]. Prototyping can enable the consideration of this context – which can relate to, e.g., physical, social, cultural, or organizational environments and influences [253] – and incorporate these crucial contextual components from early development phases.

In Section 2.3 we provide a concise overview of the notion and concepts of context and prototyping in the HCI domain, followed by a comprehensive overview of suitable methods for context-based interface prototyping of HMIs vor AVs.

1.4 Motivation and Research Questions

To enfold their potential for private and public transportation, AVs need to tackle technical and psychological challenges [132, 195]. Their adoption and their success highly depends on future users' acceptance of the technology. Challenges that might hinder their acceptance (Section 2.2.1) need to be overcome first – starting already in early development phases of new systems. As a human driver will no longer be required onboard an AV, future users will solely interact with the AI-infused systems via digital interfaces. Consequently, the interaction and communication between humans and AVs and the resulting UX are key determinants of the technology's success. Focusing on internal communication, we derived the following general research question (RQ) as a starting point for this thesis:

RQ1 How might we design suitable HMIs for (internal) human-AV interaction that counteract acceptance challenges?

Methodological frameworks such as *human-centered design* [125], *research in and through design* [59], and *collaborative UX design* [242] can serve as a suitable foundation to address RQ1. In all of them, the consideration of contextual requirements – which can be regarded as highly dynamic and complex in the automotive or (more general in the) mobility domain – plays an essential role in achieving suitable concepts for human-machine interaction (see also [143]). Here we face the following problem: AVs with SAE levels 4 and 5 driving capabilities are still under development and currently only available in specific test areas often characterized by tight restrictions (e.g., speed limits, specific test scenarios, and limited operation times). Actually, when we started working on this project, no driverless vehicles with speeds over 30 km/h were available for the public. While recent developments paved the way for publicly accessible tests of level 4 AVs without the requirements for safety drivers (e.g., [137, 206, 266]), conducting empirical research still remains limited. However, there are continuing hurdles from both technical and acceptance-related perspectives to be addressed [132, 195].

In order to facilitate public adoption, system designers need to solve human factors and acceptance challenges before AVs arrive for general public use. To do this, we need to simplify access to the context of driverless AV rides. Thus, we need to find ways to create flexible but standardizable environments while considering the potentially complex nature and dynamics of human-AV interaction in the design process. Consequently, we derived our second RQ:

RQ2 How might we prototype and evaluate (internal) human-AV interactions early considering their dynamic context?

As evaluations (e.g., usability testings) are affected by the environment in which they are conducted [253], we consider answering this question especially relevant for early development phases. In these, information on user acceptance and the likelihood of success of a system is of most significant value [60]. Since they are strongly related in terms of the context of road traffic, a good starting point to investigate RQ2 can be previous works on prototypes and evaluations of advanced driver assistance systems and driving simulations.

However, we need to note that context might not only comprise physical aspects, e.g., audio-visual impressions of the respective environment, but might also include social aspects like the presence of other people and a user's relationship with them [143]. For instance, this becomes quite important in terms of shared AVs where other travelers might influence passengers' experiences (see Chapter 3) as, e.g., the presence of others might induce stress on people [79].

Consequently, designing for such incorporates considering both the physical and the social context. Using contextualized setups, i.e., context-based prototyping [85, 117, 118], enables researchers and practitioners to do this from early development phases. This contemplation serves as a key motivation for this doctoral research. The preceding section will outline how we approach this and the two general research questions.

1.5 Methodology and Thesis Outline

At its core, this dissertation investigates the interaction between humans and AVs. Therefore, it is situated in the field of human-computer interaction (HCI). HCI is an interdisciplinary research field investigating how humans communicate and interact with computers, including information systems, the internet, or robots [9]. It draws its influences from "social, behavioral, and information science, as well as [... from] computer science and electrical engineering" [9, p. xxxiii]. Originating from a focus on user interface (UI) design, HCI research inflated to any area where computing technology affects the lives of people [9].

To answer the laid out research questions, this work makes use of established HCI frameworks. First, we adopted a *research in and through design* approach as proposed by Dalsgaard [59]. Furthermore, we embraced the *human-centered design* process as described in the ISO 9241-210 ([125]; Section 2.2.2) and combined it with the *collaborative UX design* methods conflated by Steimle and Wallach [242].

The *research in and through design* approach describes "research that (1) is directed at improving the understanding and practice of interaction design and thus includes inquiries into the design process itself, and (2) employs the researchers' involvement in design experiments as a key catalyst for knowledge generation" [59, p. 23]. It can be regarded as a combination of two approaches articulated by Ludvigsen [158], namely *research through design* and *research in design*. While research through design describes research that addresses a specific research question by using design to explore it iteratively and constructively, research in design focuses on the process, and associated events [59]. Dalsgaard [59] argues that these approaches are "not mutually exclusive" in interaction design research but, in fact, might overlap in practice, meaning that the process and the design product mutually affect each other. Research in and through design especially considers the resulting interrelations and consequent iterations [59]. In line with this understanding, we adopted a research in and through design approach to investigate the two primary research questions laid out above.

This dissertation is founded on academic papers achieved during the doctoral research. A complete list of the author's publications, including notes where they are used within this dissertation, can be found in the thesis' preface on page ix. Below, we provide an overview of the thesis' structure which is also illustrated in Figure 1.2.

Chapter 2 lays out the theoretical background and presents relevant related work in human-AV interaction and context-based prototyping. We discuss the acceptance

1 INTRODUCTION
Introducing and motivating the topic of this thesis with regard to the challenges of the current era of vehicle automation and the potential of context-based prototyping.

2 BACKGROUND & RELATED WORK
Providing a comprehensive overview of human interaction with AVs, associated acceptance challenges, and context-based prototyping methods.

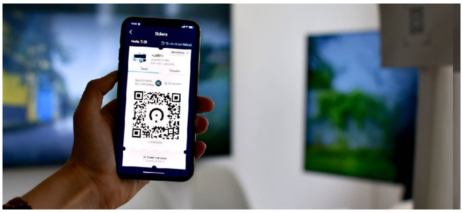
RESEARCH IN AND THROUGH DESIGN

RQ1 HMI DESIGN
How might we design suitable HMIs for (internal) human-AV interaction that counteract acceptance challenges?

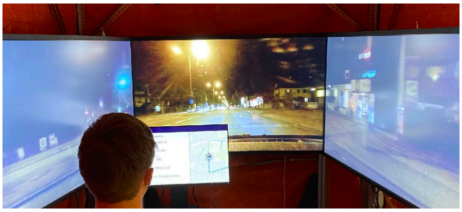
RQ2 PROTOTYPING
How might we prototype and evaluate (internal) human-AV interactions early, considering their dynamic context?

3 IMMERSIVE VIDEO-BASED SIMULATION
Presenting a cost-effective (shared) AV simulator setup using real-world videos with findings from an expert study (N = 9) and a user study (N = 31); Investigating social context simulation with actors.

4 WIZARD-OF-OZ & OBJECT DETECTION
Combining wizard-of-oz with real-time object detection for an on-road user study (N = 30) to explore how transparent (AR-based) system feedback can counteract acceptance challenges.



5 MOBILE INTERACTION WITH AMoD
Exploring two approaches for mobile interaction with AVs / AMoD (chatbots and 'classic' graphical UIs) with expert studies (N = 11) and a simulator study (N = 34).



6 INFORMATION DEMANDS AT DAY & NIGHT
Investigating the influence of day and night on the need for information about fellow travelers in shared AMoD systems with a simulator study (N = 24).

7 CONCLUSION
Summarizing the conducted work, its contributions, and answering the RQs. Laying the basis for a context-based prototyping framework and future work.

Figure 1.2 – Thesis outline.

challenges of AVs and identify suitable prototyping methods. We also provide a comprehensive overview with a qualitative semi-systematic literature review [236] and a discussion of practical considerations and applicable methods to consider physical and social contexts in AV HMI design. Together with this introduction, Chapter 2 provides the common ground, initial answers to RQ1 and RQ2, and research directions for the empirical research presented in Chapters 3 – 6 which are each based on individual publications: Chapter 3 [85]; Chapter 4 [88]; Chapter 5 [86]; Chapter 6 [87]. Therefore, Chapters 1 and 2 include parts of the introductions and related work sections presented in the publications. To avoid repeating those, Chapters 3 – 6 focus on the user studies conducted.

The first part of the reported empirical research, i.e., Chapters 3 and 4, focuses on the investigation of methods and setups for physical and social context-based prototyping of human-AV interactions (RQ2). Chapter 3 presents a cost-effective setup for immersive video-based simulation of AVs using real-world videos and a CAVE-like [55] environment. We evaluate the simulation method with an expert study ($N = 9$) and a user study ($N = 31$). In Chapter 4, we created a straightforward video-based Wizard-of-Oz (WoOz) [23] vehicle inspired by the setups of Karjanto et al. [131] and Detjen et al. [64]. We combine the setup with real-time object detection to explore how transparency in information about AVs' reasoning can counteract acceptance challenges in an on-road user study ($N = 30$).

The second part of the empirical research focuses on the design of appropriate HMIs for human-AV interaction (RQ1). Chapter 5 explores two approaches for mobile interaction with AVs and AMoD systems – namely, chatbots and 'classic' graphical UIs – with expert studies ($N = 11$) and a simulator user study ($N = 34$). In Chapter 6, we investigate the influence of day and night on the need for information about fellow travelers in shared AMoD systems in an immersive video-based simulator study ($N = 24$). Conforming to the adopted research in and through design approach, the two parts and RQs interrelate.

In Chapter 7, we summarize the conducted work and its contributions, and answer the RQs. Furthermore, we pave the way toward a context-based prototyping framework and derive potential directions for future work.

While the area of context-based prototyping methods is vast, this work puts particular emphasis on straightforward, often video-based approaches. In terms of designing human-AV interaction, we will mainly focus on explicit interaction in vehicles, i.e., internal communication. Nonetheless, most discussed and presented prototyping methods may also be applied to investigating implicit interaction and external communication.

2

Background and Related Work

Driverless AVs are increasingly developed, deployed, and publicly tested [181, 206, 264]. Before their broad introduction, however, challenges encompassing not only technological hurdles but also aspects such as ethical considerations, traffic and infrastructure management, policies, liability questions, public acceptance, and HCI need to be addressed [132, 161, 195]. Situated within this broader context, the dissertation focuses on the HCI aspect and approaches the mentioned challenges by informing and advancing the human-centered development of suitable human-AV interactions.

This chapter provides an overview of the theoretical background and relevant related work. We start with an introduction to the AV domain, automation levels, the technological challenges of AVs on a general level, and anticipated implementation scenarios of AVs (Section 2.1). In the second step, we gather a deep understanding of human-AV interactions (Section 2.2). Here we will dive into human factors and acceptance challenges, human-centered interaction design, HMI concepts, and popular evaluation methods and metrics. Building on this background, we will then concentrate on the theoretical and practical heart of this doctoral work: context-based prototyping (Section 2.3). We look at the notion of context from a general perspective, particularly in HCI, and discuss prototyping as a means to consider the dynamic and complex nature of context in designing suitable HMIs. We provide a semi-systematic review of related work on theoretical and practical prototyping considerations as well

THIS CHAPTER IS BASED ON THE FOLLOWING PREVIOUS PUBLICATIONS. PARTS OF THIS CHAPTER WERE PUBLISHED PREVIOUSLY AS PART OF THEM.

[85] **Lukas A. Flohr**, Dominik Janetzko, Dieter P. Wallach, Sebastian C. Scholz, and Antonio Krüger. 2020. Context-Based Interface Prototyping and Evaluation for (Shared) Autonomous Vehicles Using a Lightweight Immersive Video-Based Simulator. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1379–1390. `https://doi.org/10.1145/3357236.3395468`

[86] **Lukas A. Flohr**, Sofie Kalinke, Antonio Krüger, and Dieter P. Wallach. 2021. Chat or Tap? – Comparing Chatbots with 'Classic' Graphical User Interfaces for Mobile Interaction with Autonomous Mobility-on-Demand Systems. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. `https://doi.org/10.1145/3447526.3472036`

[87] **Lukas A. Flohr***, Martina Schuß*, Dieter P. Wallach, Antonio Krüger, and Andreas Riener. 2023. Designing for Passengers' Information Needs on Fellow Travelers: A Comparison of Day and Night Rides in Shared Automated Vehicles. *arXiv* 2308.02616 (aug 2023), 22 pages. `https://doi.org/10.48550/arXiv.2308.02616`.
**Note: both authors contributed equally to this paper.*

[88] **Lukas A. Flohr**, Joseph S. Valiyaveettil, Antonio Krüger, and Dieter P. Wallach. 2023. Prototyping Autonomous Vehicle Windshields with AR and Real-Time Object Detection Visualization: An On-Road Wizard-of-Oz Study. *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. `https://doi.org/10.1145/3563657.3596051`

[89] **Lukas A. Flohr** and Dieter P. Wallach. 2023. The Value of Context-Based Interface Prototyping for the Autonomous Vehicle Domain: A Method Overview. *Multimodal Technologies and Interaction 7*, 4 (2023), 1–17. `https://doi.org/10.3390/mti7010004`

[260] Dieter P. Wallach, **Lukas A. Flohr**, and Annika Kaltenhauser. 2020. Beyond the Buzzwords: On the Perspective of AI in UX and Vice Versa. In *Proceedings of the 1st International Conference on Artificial Intelligence in HCI, Held as Part of the 22nd International Conference on Human-Computer Interaction* (Copenhagen, Denmark) *(HCII '20)*. Springer International Publishing, 146–166. `https://doi.org/10.1007/978-3-030-50334-5_10`

as on methods for context-based prototyping of in-vehicle interactions (Section 2.4). The latter include lab-based mock-ups, and (VR-, video-, sound-based) simulations, wizard-of-oz setups, experimental vehicles, props, and social context simulations. Finally, we summarize the chapter as a foundation for the subsequent empirical research presented in Chapters 3 – 6.

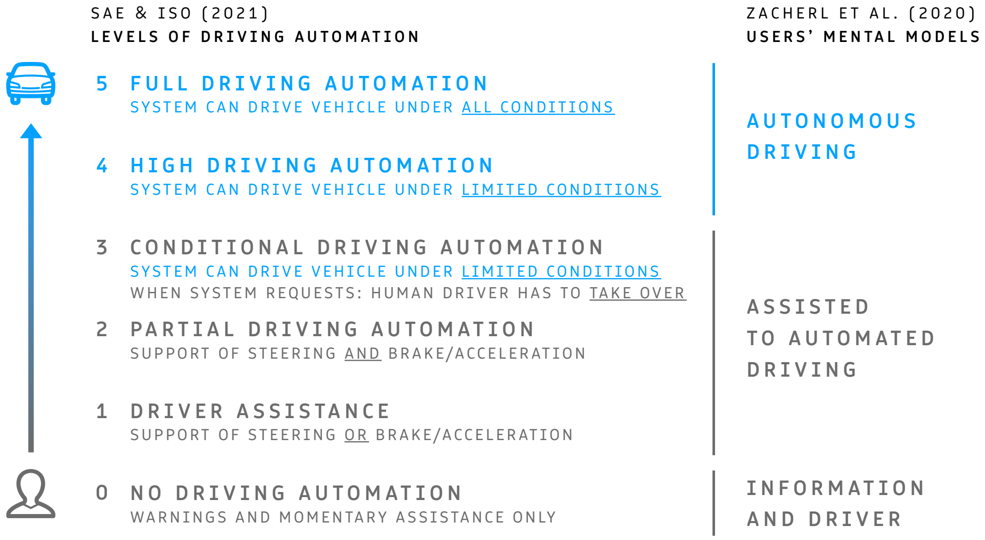


Figure 2.1 – Levels of driving automation according to ISO and SAE's J3016 [210] clustered with users' mental models identified by Zacherl et al. [278].

2.1 Autonomous Vehicles (AVs)

To describe the degree of automation in road vehicles, the Society of Automotive Engineers (SAE) [210] classifies six levels from 0 to 5 (Figure 2.1): 0) No Driving Automation, 1) Driver Assistance, 2) Partial Driving Automation, 3) Conditional Driving Automation, 4) High Driving Automation and 5) Full Driving Automation. In SAE levels 0 – 3, a human driver is still required – either to perform (parts of) the primary driving tasks (i.e., steering and acceleration) continuously (levels 0 and 1), to monitor the automation continuously (level 2), or to take over control when the automation reaches its limitations or fails (level 3). This is, however, different for levels 4 and 5 [210]. Vehicles with levels 4 and 5 systems can handle all traffic situations in their operational design domain and do not require a human driver [210]. SAE level 4 systems can drive the vehicle under particular conditions, i.e., in a specific operational design domain (e.g., only in a particular area) [210]. In contrast, level 5 systems can drive the vehicle under all conditions [210].

Within this doctoral thesis, we use the term automated vehicles as the general term to refer to vehicular systems with SAE levels 1 to 5 driving capabilities. Considering users' mental models [278], however, we use the term *autonomous vehicles (AVs)* to explicitly refer to (self-driving, driverless) vehicles with SAE levels 4 or 5 driving capabilities (Figure 2.1).

Vehicle automation and AVs, in particular, rely on processing enormous amounts of data (e.g., provided by sensors such as cameras). We then give a concise introduction to the technological background of AV information processing. Afterward, this section gives an overview of AV implementation scenarios. In the latter, we discuss the benefits and downsides of different manifestations and outline relevant ongoing (test) operations.

2.1.1 Information Processing and Technological Challenges

Similar to humans, automated vehicles need to process contextual information in order to navigate their way through traffic. The process can be divided into three main phases [67]: 1) machine perception, 2) situation comprehension and action planning, 3) trajectory planning and vehicle control (Figure 2.2). For machine perception, sensors like cameras, radars, lidars, or ultrasonic are used [277]. Those are essential for the vehicle's "perception of surroundings, localization and mapping, and vehicle state control" [277, p. 28]. Generally, AVs' technical sensors can be compared to the human senses – particularly sight and hearing. AVs use the collected sensor information in combination with their own data (e.g., current velocity), global navigation satellite systems (GNSS), and high-resolution maps. This enables ego-localization and the creation of an environmental model which serves as the basis for situation comprehension and planning. AVs often use a combination of (multi-modal) sensors to create more precise models and to increase safety and reliability through redundant sensors [277]. The process of merging information from several sensors and data sources is called sensor fusion or – more general – data fusion [257]. A crucial prerequisite for successful sensor fusion is their calibration [277]. With a remarkable resemblance to human situation awareness (Section 2.2.1; [76]), AVs' situation comprehension uses the perceived information to detect the situation and to predict future states on whose basis (driving) actions and trajectories are planned [67]. Here, AVs use technologies like computer vision and image processing (e.g., to estimate the dynamic trajectories of vulnerable road users such as pedestrians and cyclists based on a stereo vision camera [84, 140]), and AI-based algorithms. Finally, the planning is used to execute driving actions and maneuvers, i.e., vehicle control.

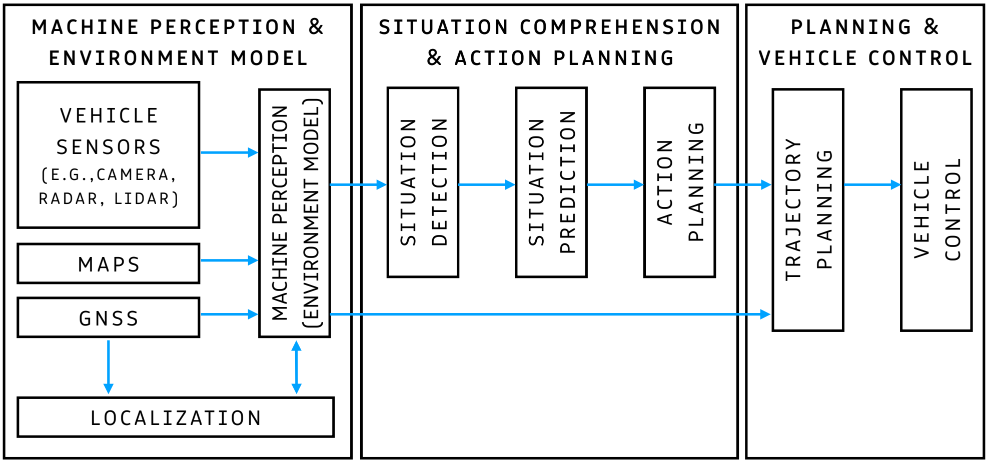


Figure 2.2 – Information processing of automated and autonomous vehicles. Illustration based on the original diagram of Dietmayer [67].

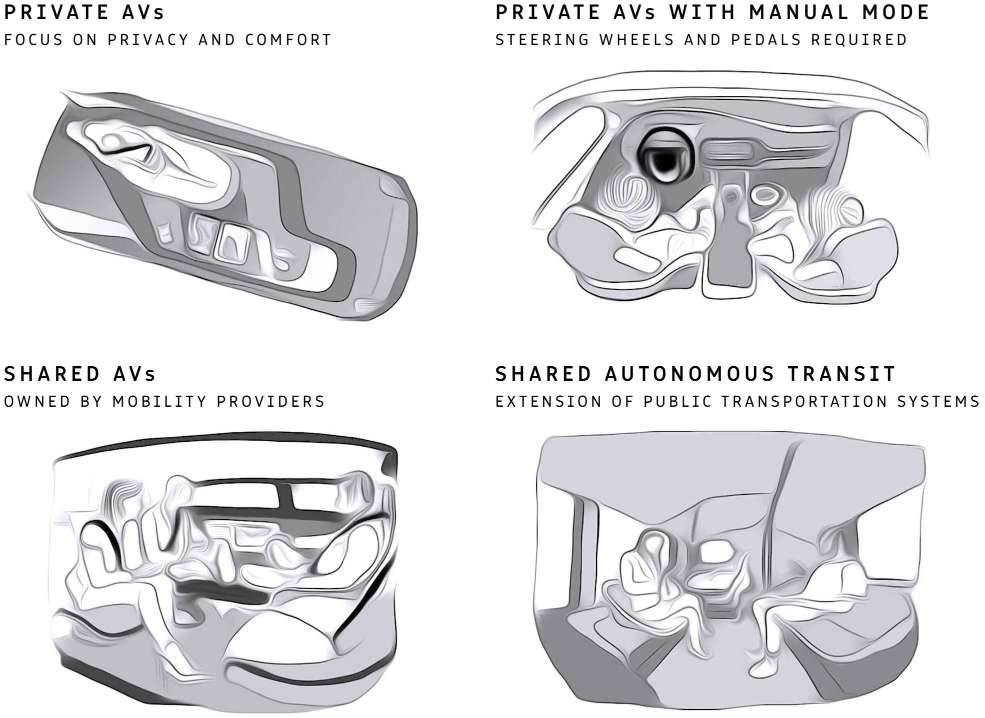
Today's automated vehicles and AVs are quite efficient in handling traffic situations with "lower" complexity (e.g., driving on a highway with a maximum speed of 60 km/h [166]), i.e., situations with clear rules, low speeds, and a limited number of other road users that need to be considered. However, complex situations with multiple agents, such as intersections in urban environments with several other vehicles and vulnerable road users, ambiguous traffic situations, and adverse weather and environmental conditions (e.g., heavy rain or fog), as well as the correct interpretation of (traffic) rules and maneuvers pose crucial challenges. In urban environments, surrounding objects with switching dynamics [140] are critical. For instance, pedestrians can rapidly change their dynamic (e.g., from standing to walking in a specific direction), resulting in fast changes in their (expected) trajectory [140]. This can lead to unforeseen and probably safety-critical events. Besides handling objects with fast switching dynamics (even at low speeds), the detection and behavior estimation of objects and obstacles at high speeds and long distances poses further problems [161]. Communication and cooperation of (automated) vehicles and other road users, as well as with the infrastructure, are discussed as options to overcome this hurdle [161]. Further improvements in the (sensor-based) perception and data fusion are expected to provide a reliable basis for AVs and enable safer and more sustainable mobility through automation [257]. When AVs can handle theses phases and the corresponding processes in their operational design domain, they will likely become continuously implemented in our (urban) environments.

2.1.2 AV Implementation Scenarios

AVs could manifest in our communities in many ways depending on factors such as environmental context, (other) infrastructure, ownership, user groups, and associated requirements. Based on Litman's [157] regularly updated AV implementation predictions, Schuß, Wintersberger, and Riener [227] derived four AV scenarios (Figure 2.3) differing in terms of ownership (private, shared, public) and degree of user control over a vehicle's driving capabilities. Below, we render the scenario descriptions of Schuß et al. [227] with some (minor) adaptions and an adapted wording considering the connotations used within this thesis (Section 2.1).

1. *Private AVs.* The vehicle is owned by a private person who is only a passive passenger without controls for manual driving but with many opportunities for non-driving related activities such as sleeping or working. A privately owned AV can be automatically rerouted by traffic management systems. For the user (and owner), this scenario offers the most in terms of comfort and privacy but is also the most expensive one.
2. *Private AVs with manual mode.* In contrast to the other scenarios, the private owner can choose whether to drive autonomously or manually (e.g., for the joy of driving). Consequently, typical driving controls like a steering wheel and pedals are still available (on request) and affect the interior design. Especially in manual driving mode, the control of traffic management systems is limited.
3. *Shared AVs / AMoD.* Shared AVs are owned and operated by mobility service providers. Passengers can book rides for a single person or groups on-demand. Rides may be shared with other (potentially unknown) passengers or booked as private rides (probably more expensive). Traffic management systems are planning and controlling the rides according to this demand.
4. *Shared autonomous transit.* Similar to shared AVs, shared autonomous transit vehicles are owned by mobility service providers and controlled by intelligent traffic management systems. The vehicles serve as an extension of (existing) public transportation systems and have the potential to cover first- and last-mile trips. From our perspective, this scenario can be considered a (sub-)variant of scenario 3 (shared AVs/AMoD).

In a qualitative user study, Schuß et al. [227] investigated the four implementation scenarios and considered shared mobility concepts most promising to contribute to a



PRIVATE AVs
FOCUS ON PRIVACY AND COMFORT

PRIVATE AVs WITH MANUAL MODE
STEERING WHEELS AND PEDALS REQUIRED

SHARED AVs
OWNED BY MOBILITY PROVIDERS

SHARED AUTONOMOUS TRANSIT
EXTENSION OF PUBLIC TRANSPORTATION SYSTEMS

Figure 2.3 – AV implementation scenarios described by Schuß, Wintersberger, and Riener [227] with adapted wording. Adapted original illustrations by Schuß et al. [227].

more livable and sustainable future. However, their results reveal opposing factors affecting people’s attitudes toward the (shared) scenarios. We will now look at the benefits and downsides of shared AV scenarios identified by related work compared to scenarios with private ownership. In a next step, we proceed with an overview of today’s public test operations of (shared) AVs.

Shared AVs and Autonomous Mobility-on-Demand (AMoD)

In general, the significance of shared mobility modes will continue to grow [56, 230]. Combining on-demand ride-sharing with AVs leads to the creation of (shared) AMoD [191] systems. AMoD can be regarded as a so-called Mobility as a Service (MaaS) solution in which users’ mobility needs are delivered through a user interface by a service provider [114, 128]. The vision of MaaS is to integrate multiple modes of transport into one cooperative ecosystem [114, 128]. In combination with (existing) public transport systems, (shared) AVs will particularly cover first- and last-mile trips of multi- and intermodal travel journeys (i.e., shared autonomous transit; Figure 2.3).

On the one hand, people see the environmental and societal benefits of shared AV rides and are open toward using it [227]. To achieve their potential, shared AVs need to provide high flexibility for passengers and enable both pre-planning of journeys as well as spontaneous trips [227] and spontaneous change of plans (Chapter 5, [86]). On the other hand, emotional factors and security concerns – especially expressed by female participants – seem to negatively affect this openness [227]. The most critical concerns are evoked by the fact that rides would need to be shared with strangers, which was found in a variety of recent studies ([214, 228, 227, 194, 213, 212], Section 2.2.1). Furthermore, the willingness to share an AV with other passengers is impaired by user uncertainty and a prevailing preference for privately owned vehicles [45]. In line with this, Pakusch et al. [188] underline the reluctance to switch from private rides to shared ones, predicting that private AVs will dominate the future of automated driving. Mapping these results to the context of shared AVs, Lavieri and Bhat [151] found that people are willing to pay extra fees for trips in shared AVs when only the vehicle, but not the trip, is shared with others. Their study indicates that privacy and security concerns might prevent participants from opting for sharing rides with strangers — affecting commuting trips to a minor extent compared to rides for pleasure purposes [151]. In general, demographic factors such as gender [211] and age [146] were revealed as predictors for the adoption of shared AVs with young men as the group with the highest openness towards using them.

Public AV (Test) Operations

Within the past decade, much data was collected in pilot and prototype operations in several research projects across the globe, e.g., [104, 181]. As a pioneer in implementing automated vehicles in urban areas, Waymo (Alphabet) began to test robotic rides in 2020 in Downtown Phoenix [265, 248]. While the vehicles actually were doing (most of) the driving, there was still a safety driver – or, as the company called it, an "autonomous specialist" [265] – accompanying the ride who could intervene if a problem occurs. Other companies, such as Cruise [206] or Baidu [137], also offered similar technologies and settings. In July 2021, Germany paved the way for the next step toward actual AVs and allowed – as the world's first state – driverless rides on selected public roads (with prior approval) with the release of the 'Act on Autonomous

Driving' [81]. Other countries follow(ed). For instance, in the mid of 2022, Californian regulators allowed actual driverless taxi services to operate in San Francisco [248]. While initially restricted to less crowded places and times (10 pm to 6 am), a safety driver was also no longer required from a legal point of view [248]. At the end of 2022 actual driverless rides became available with Cruise (General Motors) expanding its test operation to daytime rides in multiple cities [206] and Waymo now offering autonomous taxi rides availble 24/7, e.g., to connect Phoenix Sky Harbor International Airport and Downtown Phoenix [266]. Other major players in the mobility sector, such as Uber, are currently teaming up for long-term collaborations with technology providers like Motional working on driverless taxi concepts for the next decade [71]. This results in the increasing introduction of AVs in people's everyday environment, eventually allowing them to use consecutive driverless services and interact with the intelligent vehicles.

2.2 Human-AV Interaction

Interacting with autonomous systems such as AVs is a complex and interconnected matter. Yet, how people interact with systems and devices is essential for their success [218]. Prospective AV passengers will have to accept the unfamiliar – and potentially awkward experience – of driverless rides while being exposed to the decisions and actions of AI-powered systems. Lacher et al. [149] ascertain that a clear understanding of people, systems, and their interaction in a particular environment is required to establish trust in autonomous systems. In this sense, understanding human factors and acceptance challenges associated with vehicle automation and in particular with AVs is crucial. Consequently, complying with future users' requirements and counteracting those challenges is the key to successfully introducing and adopting the technology.

To counteract aforementioned acceptance and human factors challenges, human-centered design [124, 125] and human-centered AI [234, 204] can provide adequate frameworks for the suitable design and development of human-AV interaction and respective human-machine interfaces (HMIs). The following sections provide an overview of relevant aspects for designing suitable human-AV interactions and conclude with evaluation methods and metrics.

2.2.1 Human Factors and Acceptance Challenges

AVs face significant challenges from a human factors perspective and in terms of user acceptance. Besides individual (demographic) aspects of prospective AV users, the most relevant factors and challenges discussed in related work (e.g., [132, 195, 41, 181]) can be allocated to the topics overall acceptance, trust, situation awareness, safety and security, privacy, and UX. This section provides a comprehensive understanding of these factors and the connected theoretical basis for designing and prototyping suitable human-AV interactions.

Acceptance

The International Encyclopedia of Ergonomics and Human Factors defines user acceptance as "the demonstrable willingness within a user group to employ information technology for the tasks it is designed to support" [68]. In general, acceptance research is concerned with "understanding the factors influencing the adoption of technologies [. . .] by users who have some degree of choice" [68]. To better understand acceptance and its related factors, the technology acceptance model (TAM; [60]) can provide a starting point. Davis [60] developed the original version of the TAM with the objectives to "improve our understanding of acceptance processes" and provide the basis for "practical user acceptance testing". It postulates that users' attitude toward using a system is a major determinant of the actual use of a system or, respectively, users' intention to use a system [60]. Users' attitude toward using is causally influenced by its perceived usefulness (PU) and the perceived ease of use (PE), while PE also affects PU [60]. Both PU and PE are directly influenced by external (contextual) variables like system characteristics and design features [60]. Venkatesh and Davis [258] extended the model to the so-referred "TAM2" by differentiating the external variables associated with social influences as well as cognitive processes.

Since the TAM, in general, can be considered a rather abstract and high-level concept, it was often adapted or extended to better fit domain-specific requirements. This is also the case for the automotive domain. For instance, Chen [41] proposes an extension of the TAM with a focus on autonomous shuttle services (Figure 2.4). Based on a study sample of 700 participants that experienced autonomous shuttle test rides in an urban context, they conclude that besides peoples' attitudes, their perceived enjoyment directly affects the intention to use AVs [41]. Furthermore, in addition to PE and PU, passengers' trust in the system affects their attitude toward the technology

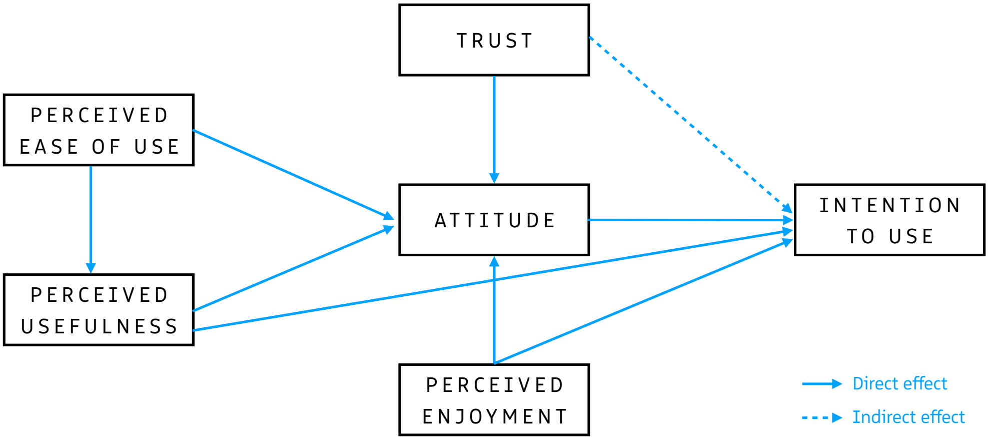


Figure 2.4 – Chen's [41] extension of the TAM. According to Chen, people's intention to use AVs and their attitude depends not only on PE and PU but also on perceived enjoyment and (indirectly) on their trust. Adapted illustration based on the original diagram of Chen [41].

[41]. Similar results were derived by Choi and Ji [43], who identify PU and trust as major determinants of peoples intention to use AVs based on an online survey with 552 drivers. Taking into account the findings of previous studies, Choi and Ji further propose to add perceived risk and driving-related personality factors to the TAM [43]. However, they did not find a significant correlation between perceived risks with participants' intention to use, but they found a causal effect of people's trust on perceived safety [43].

Based on a questionnaire study with 384 participants who experienced automated shuttle rides on a campus in Berlin-Schöneberg, Nordhoff et al. [181] use a principal component analysis to derive three acceptance components with positive relations: intention to use, shuttle and service characteristics, and shuttle effectiveness. In accordance with that, Kaur and Rampersad [132] found people's performance expectancy as a significant determinant for AV adoption. In line with Chen [41] and Choi and Ji [43], they also identified trust as a major determinant for AV acceptance [132].

Trust

Trust can be defined as "a belief that something is expected to be reliable, good and effective" [149, p. 43] and as a mental state people have based on their expectations and perceptions [149]. Consequently, trust can also change with experience [174].

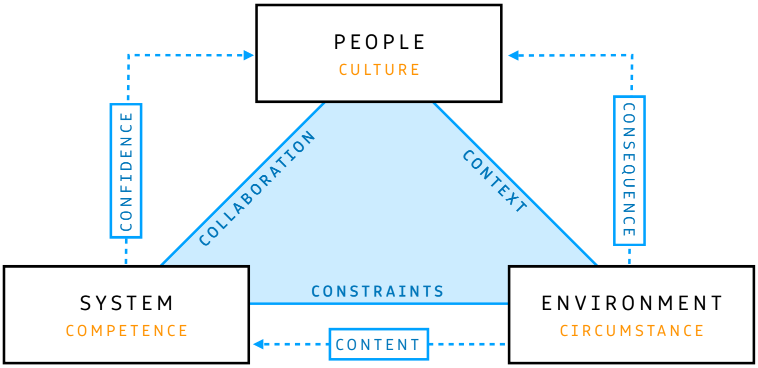


Figure 2.5 – Framework for discussing trust in autonomous systems by Lacher et. al [149]. Illustration based on the original diagram of Lacher et al. [149].

Many problems associated with automated systems can be assigned to people failing to interact with and rely on the automated system in an appropriate manner [152]. Lacher et al. [149] point out that trust depends on multiple parties depending on various factors associated with certain roles in autonomous systems. They conclude that a precise understanding of users, systems, and their environments is required to establish trust and propose a corresponding framework to discuss trust in autonomous systems [149]. Their framework (Figure 2.5) illustrates the interrelations of these aspects and that users' trust does not only depend on single components like the competence (or performance) of a certain system (e.g., an autonomous taxi), but also on (trust in) other people and institutions (e.g., vehicle manufacturers, city councils) as well as environmental circumstances. This underlines the importance of physical and social context for system trust and acceptance. Resulting constraints and consequences need to be considered to enable suitable collaboration between people (users) and (autonomous) systems.

Lee and See [152] also describe how trust is influenced by individual, organizational, cultural, and environmental context. In order to avoid misuse and disuse of automation, they, furthermore, follow that an appropriate "calibrated" level of trust is required [152]. This means that automated systems clearly need to communicate their capabilities in order to match users' trust with actual system capabilities (trustworthiness) and consequently avoid overtrust (that might lead to misuse) and distrust (that might lead to disuse) [152]. Figure 2.6 illustrates the interrelation of trust and trustworthiness.

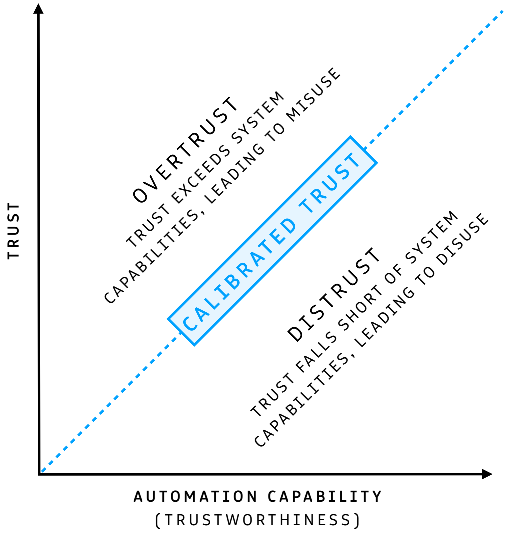


Figure 2.6 – Lee and See's [152] concept of calibrated trust, which depends on a match of users' trust with actual automation capabilities. Illustration based on the original diagram of Lee and See [152].

Since trust influences users' reliance on automation [152], it can be regarded as a crucial component for AV acceptance. In accordance with the more general perspective of Lee and See [152] and based on their embedded quantitative case study on an Australian university campus with 101 participants, Kaur and Rampersad [132] identify users' trust as the primary adoption barrier of AVs, since people need to give up control and choice to the autonomous system. In line with Lacher et al. [149], Kaur and Rampersad further conclude that the reliability of AVs and its match with users' performance expectations can be associated to their trust in AVs [132]. To increase (appropriate) trust in AVs, Eden [74] concludes that easy to understand human-AV interactions are required. Riedl [204] puts forth that transparent communication can create the basis for such an understanding and advocates for the application of a human-centered AI approach (Section 2.2.2) to achieve it. Such would also lay the basis for users' awareness of the usage situation.

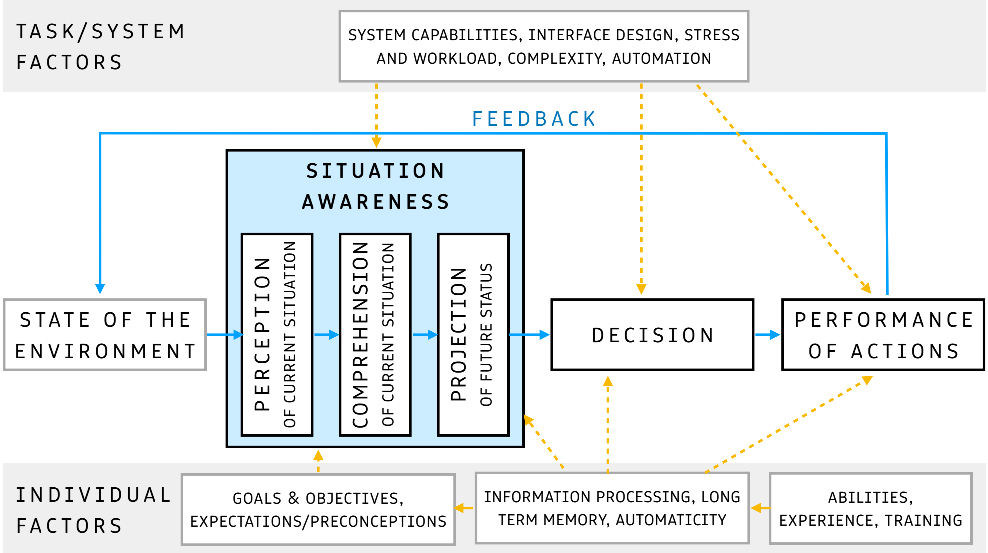


Figure 2.7 – Endsley's model of situation awareness [76]. Illustration based on the original diagram of Endsley [76].

Situation Awareness

Situation awareness describes a person's knowledge of a specific dynamic situation or environment [76]. The construct of situation awareness can be divided into three levels [76]: 1) perception of elements in the current situation, 2) comprehension of the current situation and 3) projection of future status based on this information. As illustrated in Figure 2.7, situation awareness is highly related to humans' decision-making process as well as individual-, system-, and task-based influencing factors. Endsley concludes that good situation awareness helps to make and execute appropriate decisions [76]. Automation as a system-related factor, on the other hand, decreases situation awareness by changing a users' role from active to passive [77]. In lower automation levels (i.e., SAE levels 1 – 3), insufficient situation awareness can be a critical factor and a severe safety issue with increasing automation – e.g., when a take-over request occurs due to an unexpected system failure.

In contrast to that, this is not that critical in AVs (i.e., SAE levels 4 and 5) from a safety point of view. Since, e.g., humans are not required – or not even able – to take over the driving task in AVs, one might even say that – depending on the system design – their ability to decide and execute actions is restricted in the context of AVs. However, humans are still required to have an appropriate level of trust toward autonomous systems (Section 2.2.1). As pointed out by related research (e.g., [149, 152]), environmental and contextual factors are important determinants of trust. In order to "help humans understand the AI systems" [204, p. 35], humans need to understand the context and status of the AV and are consequently also required to gain a sufficient level of situation awareness. Therefore, it remains – similar to lower automation levels – important to facilitate effective communication between humans and AVs to enable a sufficient level of situation awareness. This can manifest in HMIs providing information to increase this awareness, e.g., by providing information on the current system status, driving context (e.g., location, planned route), and explanations of the automated system's decisions, intentions, and actions (e.g., obstacles or maneuvers such as braking or turning) [82, 142, 156, 195]. Providing users with information on the surrounding elements of automated vehicles can increase users' situation awareness [156] and consequent trust in the system [271, 270, 185]. With reference to Fröhlich et al. [95], Pigeon et al. [195] point out that several modalities (e.g., visual, auditory) and formats (e.g., text, icons, augmented reality (AR)) might be used and combined to convey this information, but current research has not yet derived which might be most appropriate and comprehensible.

Safety, Security, and Privacy

Previous studies identified factors and concerns about traffic safety (i.e., the prevention of inherent threats and risks [267]), e.g., provided by the correct functioning of the automated system as well as security issues induced by the absence of a human driver [195]. Functions such as (emergency) stop buttons [182] and means of communication [195] were identified as requirements to increase passengers' (feeling of) safety and security. In contrast to safety-related aspects, security concerns refer to the fear of deliberate internal or external risks [267], i.e., intentional actions by other (human) entities. For instance, security concerns can refer to the fear of hacker attacks on the autonomous driving system [148, 132].

Furthermore, concerning shared AV scenarios, and especially when AVs become integrated into public transportation systems, rides will be shared with other, unknown people. This raises concerns among potential users about interactions with strangers without human oversight, like a driver in a conventional bus would provide [227]. For instance, people fear insults and assaults from fellow passengers [195]. Studies from related work emphasize resulting security concerns as an important issue for automated ride sharing, especially for women and particularly during the night

[187, 194, 212, 227, 228]. Here, the absence of a human driver in shared AVs was identified as a critical aspect [93, 196, 212, 227]. Lavieri and Bhat [151] found that not having a driver on board in an AV seems to be particularly problematic for Millennials and Generation Z (at the time of the study (2019): people between the ages of 18 and 34), as they see a driver as a kind of "guardian". In line with that, Biermann et al. [24] conclude an increased need for security in shared AVs.

Consequently, future passengers might accept the application of monitoring systems to prevent crime, vandalism, and also in case of health emergencies [24]. However, the use of surveillance systems such as closed circuit television (CCTV) or audio recordings of the vehicle's interior is, like the need for registration to use the resulting mobility services, regarded critically – especially regarding (personal) data protection [132, 240]. Stark et al. [240] emphasize the resulting ambivalence between security measures and privacy requirements. This matter might also affect passengers' UX.

UX and Usability

UX can be described as a *"person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service [... . It] includes all the users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors and accomplishments that occur before, during and after use"* [124, p. 7]. As a result, UX can be considered *"a consequence of the presentation, functionality, system performance, interactive behaviour, and assistive capabilities of an interactive system, both hardware and software. It is also a consequence of the user's prior experiences, attitudes, skills, habits and personality"* [124, p. 7].

UX is often inappropriately used as a synonym for the term usability. According to the ISO 9241, usability can be defined as the *"extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"* [124, p. 7]. We want to note the difference in the temporal aspect. Usability focuses on the actual usage situation, while the concept of UX includes the expectations and anticipations of a prospective users that contribute to the total of their experience with a system, product, or service – just as their experience during the actual usage situation and their retrospective considerations after use do. Consequently, we regard usability as a part of UX. Additionally, we consider the work of Hassenzahl et al. [109, 110] and Laugwitz et. al [150] fitting who differentiate two aspects of UX: 1) pragmatic quality, which is often referred to usability [109] and 2) hedonic quality, which describes humans' desires for stimulation and identity [109] and can be related to user satisfaction [150].

The usability or – more general – the UX of HMIs plays an essential role in terms of user trust and acceptance [3]. Since HMIs will provide the only touchpoints with driverless AVs and respective AMoD systems, their usability and overall UX greatly impact their acceptance. Acceptance factors such as perceived usefulness, perceived ease of use, perceived enjoyment, and intention to use (Section 2.2.1; see Chen [41]) are likely to be directly affected by AV passengers' experiences before, during, and after a ride. To positively affect these experiences, human-AV interactions need to be designed by considering humans' needs and requirements from early development phases. The human-centered design process [124] provides a framework to do so.

2.2.2 Human-Centered Design (HCD)

Human-centered design (HCD) can be described as *"an approach to interactive systems development that aims to make systems usable and useful by focusing on the users, their needs and requirements, and by applying human factors/ergonomics, and usability knowledge and techniques"* [124, p. 4]. By taking into account the characteristics and capabilities of users, it aims to enhance "effectiveness and efficiency, improves human well-being, user satisfaction, accessibility and sustainability" [124, p. 4] and to counteract "possible adverse effects of use on human health, safety and performance" [124, p. 4].

HCD Process

At its core, the HCD process described in the ISO 9241-210 [124, 125] consists of four iterative design activities which are conducted after an initial planning phase (Figure 2.8): 1) understanding and specifying the context of use, 2) specifying user requirements, 3) producing design solutions, and 4) evaluating the design solutions. If the design evaluation reveals the need for optimization, the individual activities are run through again in the form of iterations. The process is iterated until the design solution meets the identified user requirements.

In this iterative process, the evaluation is usually conducted as a formative evaluation focusing on the investigation of "which aspects of the design work well or not, and why" [130]. The key motivation of formative evaluation is to inform the design process and to improve on the current state based on the information derived [130]. On the other hand, summative evaluation is usually used to refer to evaluation activities that assess "how well a design performs" [130]. Thus, summative evaluation

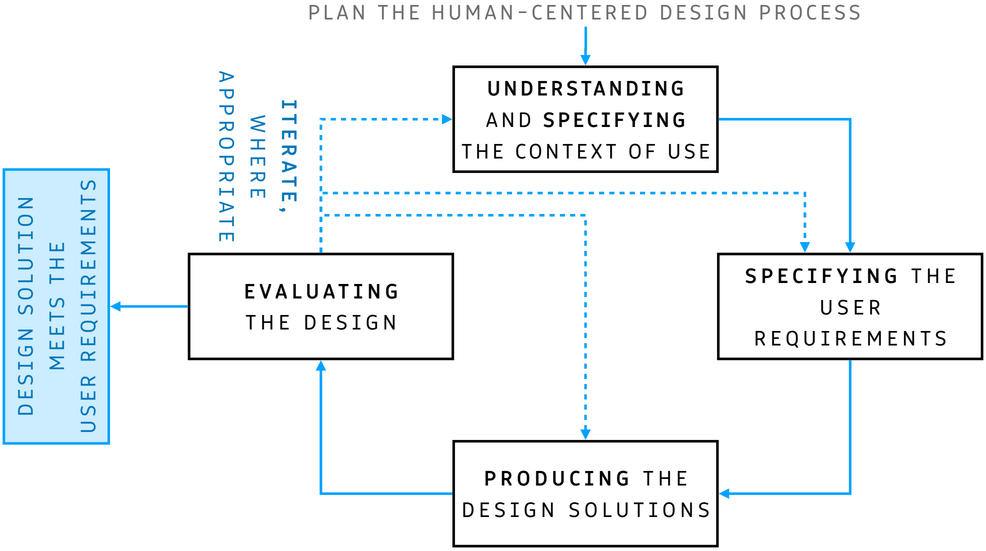


Figure 2.8 – The human-centered design process according to ISO 9241-210. Illustration based on the original diagrams of ISO 9241-210 [124, 125].

is usually conducted at the start or the end of a design/development process and aims to investigate "the overall experience of a finished product" while it is also often used to compare the results against benchmarks or competitor products [130].

Contextual aspects play an important role in HCD. We, thus, regard it essential to consider the context in all of the design activities for a valid application of the process. Regarding the AV domain, this motivates the development and application of suitable prototyping methods enabling the context-based conduct of HCD activities (Section 2.3).

Human-Centered Artificial Intelligence (HCAI)

Interacting with AVs implies interacting with systems based on artificial intelligence (AI). AI, *"broadly (and somewhat circularly) defined, is concerned with intelligent behavior in artifacts. Intelligent behavior, in turn, involves perception, reasoning, learning, communicating, and acting in complex environments. AI has one of its long-term goals in the development of machines that can do these things as well as humans can, or possibly even better"* [179, p. 1].

Considering the cited definition by Nilsson [179] from a broader human-factors-oriented view, we can see that there is a notable overlap of AI with what we call automation [189, 190, 260]. To describe the degree of automation in intelligent systems,

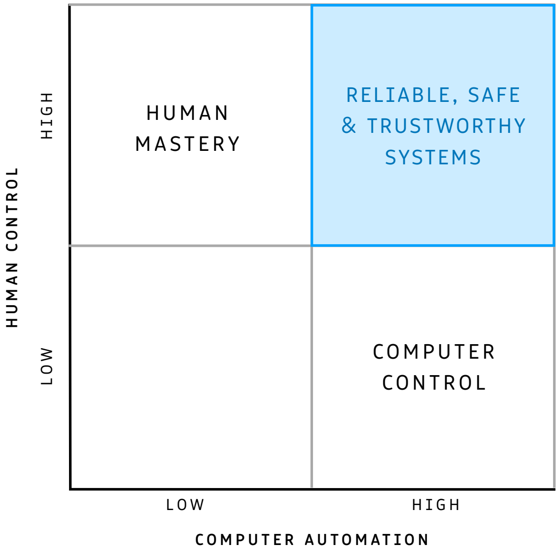


Figure 2.9 – Shneiderman's two-dimensional framework to describe human-automation interaction with the goal to achieve reliable, safe, and trustworthy systems by having high levels of human control and high levels of automation [233]. Illustration based on the original diagram of Shneiderman [233].

most frameworks use one-dimensional scales to describe the division of tasks and responsibilities between humans and machines, e.g., [83, 210]. In contrast to those, Shneiderman [233] suggests a two-dimensional framework to think about automated human-machine systems more collaboratively by simultaneously considering the dimensions: degree of human control, and degree of computer automation (Figure 2.9). Shneiderman proposes creating reliable, safe, and trustworthy systems which can be achieved by high degrees of both human control and computer automation [233]. However, suitable human-machine systems can also have low computer automation, but high human control, which would, according to Shneiderman, result in 'human mastery', or – on the contrary – high computer automation and low human control, which would result in 'computer control' [233], Figure 2.9.

As mentioned before, in intelligent systems such as AVs and AMoD, humans interact with automated systems founded on AI. Consequently, users experience AI. We agree with Cramer and Kim that "the field—where UX meets AI—is full of tensions" [51, p. 69]. Therefore, recent research suggests designing new (AI-based) systems in a human-centered way and postulating human-centered AI (HCAI). Riedl describes HCAI as the perspective to designing AI algorithms with the "awareness that they are part of a larger system consisting of humans" [204, p. 33]. Besides understanding humans, their abilities, and needs, a critical aspect of HCAI is "to help humans understand AI systems" [204, p. 35].

The approach of providing explanations as a means to make AI systems (more) comprehensible for humans is referred to as explainable artificial intelligence (XAI) [105]. Drawing upon the works of Bellotti and Edwards [16] on human considerations in context-aware systems, Gunning [105] puts forward that XAI systems should be able to explain their capabilities and understandings. Furthermore, they should be able to explain their past, present, and future actions and reveal the information basis on which they are acting [105]. Such explanations and, more general, transparent communication can provide the basis for the targeted understanding in terms of HCAI and, consequently, increase people's confidence and willingness to use these systems [204].

In line with the concepts of XAI and HCAI, Amershi et al. [6] suppose to (initially) clarify what the system can do and how well it can do that. Building on Horvitz' principles for mixed-initiative user interfaces [119], they propose design guidelines for human-AI interaction which are also concerned with (contextual) information and feedback during the interaction, error handling, and long-time interaction.

Along with the general human-centered design process [125], the discussed concepts, frameworks, and guidelines of human-centered AI provide a well-founded starting point for the creation of adequate HMIs.

2.2.3 Human-Machine Interfaces (HMIs)

For the interaction of humans and AVs, human-machine interfaces (HMIs) are required. In the HCI domain, (user / human-machine) interfaces refer to physical facilities or hardware that enable the communication between humans and machines [121]. They provide the basis for the communication between passengers and the intelligent systems since human operators (e.g., drivers) are no longer involved. In this context, the term user interface (UI) is often used as a synonym for HMI what we will also adopt in the following sections.

Interfaces can generally be divided into displays and control elements (Figure 2.10; [32]). As explained by Bubb et al. [34], displays can be described as technical elements with which humans can receive information from a machine through their senses.

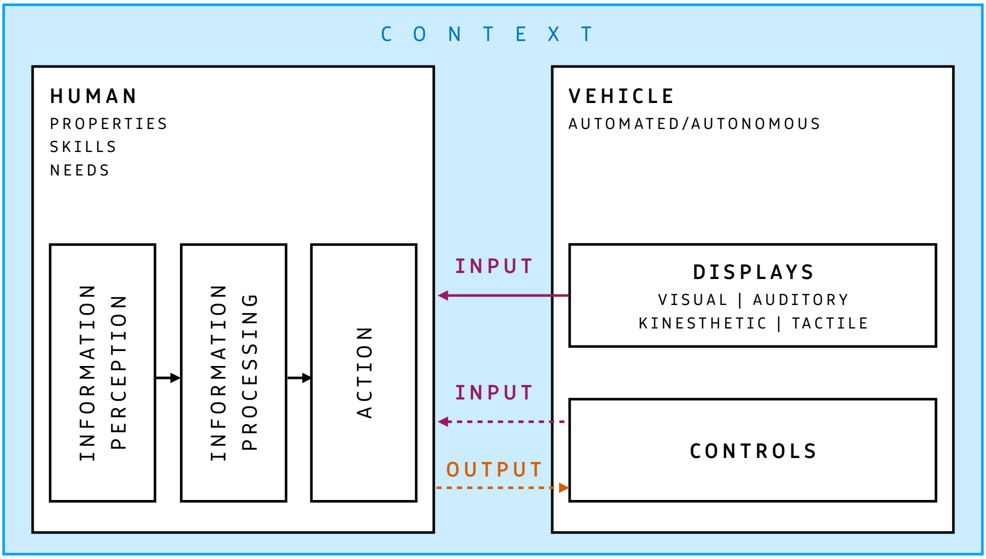


Figure 2.10 – Adapted working model of human-machine interaction in/with (automated/autonomous) vehicles. Illustration based on the original diagram of Bruder and Didier [32].

Here, all human senses are eligible, while in particular visual, auditory, and haptic displays are used primarily [34]. Controls can be described as technical facilities enabling humans to transfer information to a machine [34]. Here, extremities (arms, legs), feet, hands, or fingers can be used to actuate physical elements (e.g., buttons, pedals, levers) but also to perform gestures (e.g., on a touchscreen). Furthermore, machine sensors can detect gestures, noises, and speech (e.g., via microphones, cameras, and infrared sensors) and also serve as controls [34].

HMI Types

With the objective to formalize the various manifestations of HMI displays and controls that can be used to interact with automated vehicles, Bengler, Rettenmeier, Fritz, and Feierle [20] introduce a framework to describe the different types of HMIs, their interrelations among each other, and other influencing factors (Figure 2.11). Generally, they distinguish between internal communication (i.e., interactions inside the vehicle) and external communication (i.e., interactions outside the vehicle, e.g., with other road users). The latter is, e.g., concerned with *external HMIs (eHMI)* enabling AVs to communicate their intentions and maneuvers (e.g., stopping, accelerating, giving way) to human drivers of other vehicles or vulnerable road users [113] (e.g., pedestrians and cyclists). Based on previous work on human-human interaction, Schieben et al. [216] derive four main categories of information that AVs should communicate to other road users: its current automation status (i.e., the activated driving mode), its planned maneuvers, its perception of the environment, and its cooperation capabilities. Within the scope of this thesis, we are focusing on internal communication with AVs. Here, Bengler et al. [20] differentiate four HMI types:

- *Dynamic HMI (dHMI).*
  Refers to mostly implicit communication via the vehicle's dynamics.
- *Vehicle HMI (vHMI).*
  Provides general information on the vehicle's condition.
- *Infotainment HMI (iHMI).*
  Offers information and interactions for non-driving related activities.
- *Automation HMI (aHMI).*
  Provides information on the automation's system status including information on current and future events.

In practice, such a clear differentiation is not always feasible. Instead, HMIs can, on the one hand, combine several of the types identified by Bengler et al. [20] in a single component. On the other hand, e.g., iHMIs and aHMIs might be distributed across several locations and modalities or displayed redundantly.

Jansen et al. [126] provide a comprehensive overview of the in-vehicle design space for input and output modalities and information locations, and an extension of the previous vehicle design spaces by Detjen et al. [63] and Kern and Schmidt [133] based on an exhaustive literature review. Moreover, they highlight the importance of multimodal in-vehicle interactions [126]. Their systematic literature review reveals that the most established modalities for human-vehicle interaction are visual, auditory, kinesthetic, and tactile [126]. This is in line with the above-introduced fundamental literature of Bubb [34] and Bruder and Didier [32] (Figure 2.10).

For interacting with (shared) AVs and AMoD, already familiar HMI concepts and components like touchscreens, information displays and control buttons seem to be preferred by perspective users [24]. A preference for established locations, such as the front area, is also reflected by Jansen et al.'s review [126]. However, they also point

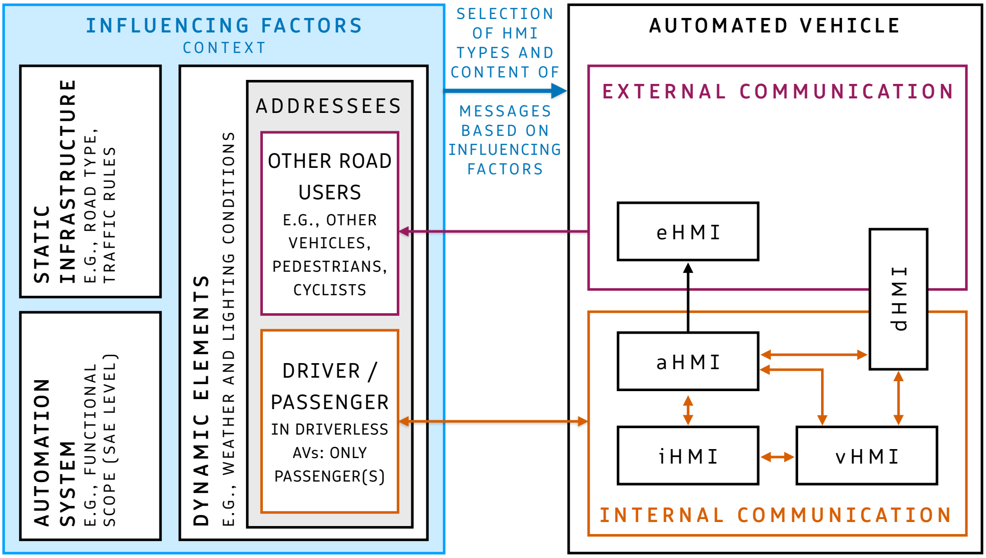


Figure 2.11 – HMI framework proposed by Bengler, Rettenmeier, Fritz, and Feierle [20] illustrating the interrelation of the HMI types among each other (right part) and with the influencing (contextual) factors (left part). Illustration based on the original diagram of Bengler et al. [20] (CC BY 4.0).

out that future automotive HMI designers may consider novel modalities such as "vestibular, thermal, olfactory, cerebral, or cardiac" [126, p. 8] which might become more approachable and usable by passive passengers in AVs.

As mentioned, AV passengers are not required to monitor the vehicle's condition and settings as a driver would need to do via a dHMI and a vHMI in vehicles with lower automation levels. In order to support passengers in understanding the AI-powered system [204] as well as its intentions and actions, transparent internal communication via iHMIs and aHMIs is, therefore, placed in focus and could be the key to high acceptance, trust, and positive UX.

Explicit vs. Implicit Interaction and the Interaction-Attention-Continuum

Previously, human drivers interacted with other traffic participants directly. With increasing automation of vehicles and in particular with the introduction of AVs, current road traffic will shift to mixed traffic [216]. This means, that there will be both, vehicles with automated driving capabilities and road users without automation, simultaneously present in the same driving context. As a result, the interaction shifts to a triad where on-board users need to interact with the vehicle automation which interacts also with other traffic participants and replaces to some part or fully (depending on the automation level) the interaction of on-board users with other road users [216].

Generally, we can distinguish explicit and implicit interactions between humans and humans as well as between humans and machines. *Explicit interaction* with machines happens when users tell a system "in a certain level of abstraction (e.g. by command-line, direct manipulation using a GUI, gesture, or speech input) what they expect the computer to do" [218, p. 192]. In contrast, *implicit interaction describes* actions performed by users "that [are] not primarily aimed to interact with a computerised system but which such a system understands as input" [218, p. 192]. For instance, pedestrians can communicate implicitly to a human driver or an AV that they have seen the vehicle and are likely to give way by looking at it and stopping at an intersection [66]. For implicit human-AV interaction, AVs need to understand the situation and the intentions of the user. Similarly to their information processing in terms of driving automation and human situation awareness, the system needs to perceive, understand, and predict the situation and the intention of the user considering both explicit and implicit communication (Sections 2.1.1, 2.2.1). Stampf et al. [239] provide a comprehensive literature overview on which states and user intentions might be detected and communicated through implicit interaction.

Beyond explicit and implicit interaction, Bakker and Niemantsverdriet [10] propose an *interaction-attention-continuum* (Figure 2.12) to describe the relationship between human attention (on a spectrum from fully focused attention to completely outside the attentional field) and corresponding interaction types. The authors identify three interaction types: focused, peripheral, and implicit interaction [10]. They suggest that interfaces should facilitate interaction at all of these varying and shifting levels to fit ubiquitous systems into people's everyday life and routines [10]. While we already discussed implicit interaction and considering that the concept of focused interaction can be roughly mapped to explicit interaction, peripheral interaction provides a third category. It holds characteristics of both explicit (intentional, direct control) and implicit interaction (subconscious; Figure 2.12). In the context of (autonomous) mobility, this may, for instance, concern scenarios occurring in routinized travels such as commutes. For example, a commuter might subconsciously count the number of stop announcements on a fixed route to know when to get off. At the same time, another passenger may be on the very same ride for the first time and is consciously focusing on the information provided by the system.

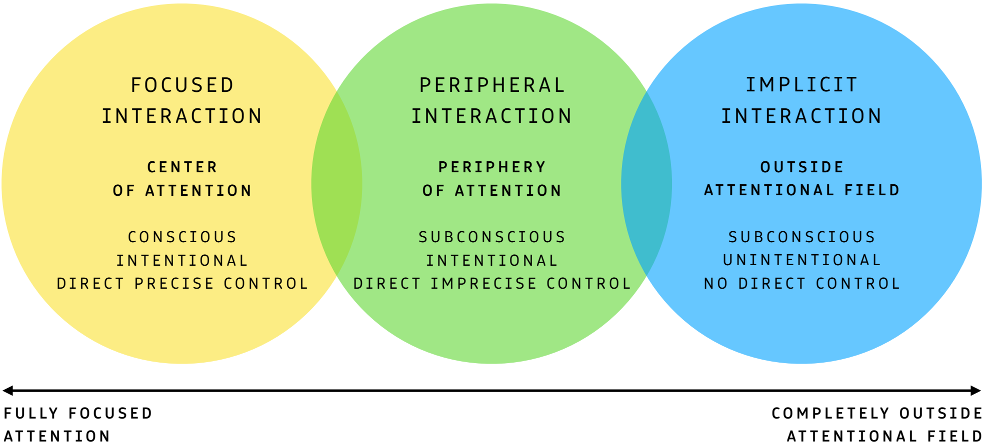


Figure 2.12 – Interaction-attention continuum proposed by Bakker and Niemantsverdriet [10]. The three general interaction types (focused, peripheral, and implicit) are mapped to the human attention spectrum. Each type is described with the corresponding level of attention and interaction characteristics. Illustration based on the original diagram of Bakker and Niemantsverdriet [10] (CC BY 4.0).

As Bakker and Niemantsverdriet [10] conclude, it is, therefore, vital for designers to consider the varying interaction types and corresponding information demands.

Information Demands

Surveys and empirical studies point out that prospective AV passengers demand information they are already used to receiving from other private and public transportation systems, such as details on the location, planned route, or upcoming stops [24, 85, 193]. Related work suggests using visual and auditory feedback about the next stop of automated shuttles [159]. Considering shared AV rides, however, it needs to be taken into account that travel information can be a private matter [30]. König et al. [139] evaluated whether information about potential co-passengers influences the acceptability of shared AMoD systems and measured how different levels of information affected participants' compensation demands. Detailed information about co-passengers proved to be beneficial [139]. Interestingly, they also found that information about men as fellow travelers resulted in higher refusal rates than information about female travelers [139]. In accordance with this observation, women seem to prefer being matched with other women to increase feelings of security [196, 228].

In line with XAI and HCAI approaches arguing for transparent communication between humans and machines (Section 2.2.2), Norman [183] already pointed out in 1989 that the "problem" of automated systems is often inappropriate feedback and interaction. Norman proposed the provision of continuous feedback on the system's system status, the anticipation of errors from both humans and machines, and the anticipation of the occurrence of the worst possible situation [183]. Recent research on automated vehicles also indicates that contextual information and information on the AI-powered systems' status, reasoning, and actions might affect the perceived safety, trust, acceptance, and UX of AV passengers [82, 142, 156, 160, 185, 270]. Based on available AV sensor data (e.g., detected objects), HMIs could provide such information to supply passengers with transparent explanations. These might be able to compensate (at least for some part of) the absence of a human driver and counteract said challenges.

Oliveira et al. [185] found in an indoor study with an experimental level 4 AV that providing transparent system information via HMIs can increase trust. In their comparison of HMI concepts, an AR-based variant received the best assessment [185]. Equivalent results have also been reported for AR interfaces in vehicles with lower levels of driving automation, e.g., [156, 271, 270]. Wintersberger et al. [271, 270], based on a simulator study, conclude that traffic augmentations can increase driver's trust in ambiguous situations (e.g., dense fog). However, not all users might want to receive such information at all times [185]. Thus, the design and amount of provided information and explanation are crucial since "more information does not necessarily lead to more trust" [160]. Similar results can be observed in other domains: e.g., Kizilcec et al. [136] found that making an algorithmic interface for peer assessment more transparent by providing explanations can increase trust but also diminish already built confidence if too much information is provided. It needs to be investigated whether these findings also apply to the context of driverless AVs in real-world driving scenarios.

Colley et al. [47] investigated the potential of semantic segmentation visualization of detected objects to increase trust and situation awareness in vehicles with highly automated driving systems (i.e., SAE level 3) by conducting two online studies. Their findings revealed the potential of AR-based visualization to increase situation awareness, but did not reveal significant effects on participants' trust [47]. In a consecutive work, Colley et al. [48] investigated AR-based visualizations of different types of information related to a level 3 automated vehicle's phase of information processing: situation perception, situation prediction, and action planning (Section 2.1.1). They conducted an online study to compare respective visualizations and their combination for each of these levels by showing participants single videos with respective augmentations. Their results show the effects of the visualization variants on participants' trust, workload, situation awareness, and perceived safety [48]. Colley et al. [48], in particular, found visualizations related to situation prediction are perceived negatively and can degrade the attributed capabilities of the automated vehicle. However, the authors conclude that visualizations can serve as a countermeasure to overtrust by educating future users (here, SAE level 3 drivers) and, consequently, help to calibrate trust. It needs to be evaluated and derived to what extent these findings apply to driverless AVs and how suitable HMIs for (in-vehicle) human-AV interaction need to be designed.

UI concepts for Shared AVs and AMoD

AMoD UIs can range from personal planning and booking applications on various devices (e.g., smartphones, tablets, desktops) to in-vehicle HMIs and terminals at mobility hubs with diverse in- and output modalities. For interacting with AVs, users seem to prefer established technologies such as smartphone booking apps, in-vehicle touchscreens and control buttons but tend to reject less familiar methods, including analogue hand gesture communication [24].

As already discussed: since there is no human driver onboard a (shared) AV, the question arises of how to communicate with the AV and the service throughout the journey – and especially during the ride – considering acceptance, privacy, and trust issues (e.g., [30, 132]). This question is inevitable with regard to occurring change of plans (CoP), e.g., the need to change the departure time or target destination of a booked trip or abort an ongoing trip. In general, the services of a human driver might be substituted by a mobile app [24] serving as a personal travel companion. Such an app could rely on a 'classic' graphical user interface (GUI; e.g., [69], Figure 2.13:i,vii), but also on a conversational user interface (CUI) (e.g., [135]). As cost-effective AMoD rides would be shared with other passengers, the interaction between users and system will include interacting in public spaces with other people present. Given that travel information can be a private matter [30], visual in- and output seems superior to other interaction modalities like, e.g., speech. As a consequence, 'classic' GUIs and chatbots (i.e., text-based CUIs) appear to be promising approaches for a mobile AMoD companion app. Below, we provide an overview of the two concepts and their advantages and disadvantages discussed in the literature.

GUIs. GUIs with touch-input can be considered as the de facto status quo of interacting with computer systems on mobile devices. This also applies for currently available ride-sharing and mobility-on-demand services like, e.g., Uber, MOIA, CleverShuttle, or Free Now. Designing for GUIs' usability and positive user experiences implies to think about navigation patterns, menu structures, and the interaction with graphical elements, editable text fields or buttons [90, 123]. Generally, GUIs can – in contrast to CUIs – easily provide an overview of a system's functionalities and scope [170, 244] and are suitable to display plenty of information [37]. Based on that, users can build a clear mental model of the system (e.g., [170]). This makes it easy for them to choose between provided options and discover the system through visual clues (e.g., [170]). As a result, GUIs can provide efficient shortcuts [245] to access specific functions (e.g., aborting an ongoing (shared) AMoD ride with a single button). Although GUIs often use established concepts, users must be able to understand the layout and visuals, their underlying logic, and the interaction concept [244, 245]. This is a major design challenge, especially concerning new usage contexts like AMoD that come with new functions and restrictions.

Chatbots. In general, "the front-end to a chatbot or virtual personal assistant" [165, p. 40] is provided by a CUI enabling users to interact with natural language and in- and output modalities like, e.g., speech, text, or touch [164, 165]. When users express their needs in their own words [244], the system needs to understand their intents [255]. In contrast to voice-based virtual assistants, chatbots are text-based and require a graphical counterpart. Their visual UI can be considered as a blank canvas providing content and features on demand [90]. Designing for chatbot usability implies providing users with the appropriate information at the right time and making good suggestions [90]. If done right, chatbots can simplify the information search process [255], especially in complex search spaces. Through the similarity with natural conversations and instant messaging apps, they can provide convenience and ease of use [255]. By taking contextual information acquired in previous conversations into account, bots can also personalize the interaction based on individual users' characteristics [255, 279] and offer context-based 'shortcuts' to step-by-step approaches commonly used by 'classic' GUIs. For example, in shared AMoD rides, it might typically not be possible to change the destination of an ongoing trip as this would also affect all other passengers' rides. However, suppose a user requests such a change of plans. In that case, a chatbot could consider the conflict with the ongoing ride, inform the user about it and directly suggest a solution (e.g., leave the current ride at a suitable location and change to another connecting shuttle). In contrast, a ‘classic’ GUI would typically offer a step-by-step strategy (e.g., abort the current trip in a first step and then book a new ride). As interactions with chatbots are currently often either productivity-oriented or relational, Følstad and Skjuve [92] suggest to integrate both forms to enhance conversational UX. The relational aspect, i.e., the creation of a natural and ‘human-like’ feeling [102] is regarded as a key challenge in designing chatbots. To achieve this, the bot’s personality needs to be clearly defined [107, 232].

State of the Art: UI Examples from the Industry

Supplementing the above-discussed theoretical and scientific related works, this section looks at the current state of the art of human-AV interaction design in the industry. Figure 2.13 provides an overview of exemplary UIs used by leading AV technology companies such as Baidu [8], Cruise [53, 54], and Waymo [197, 250]. As mentioned in Section 2.1.2, all three companies are currently testing autonomous rides. In this section, we discuss particular aspects of the used UIs in the context of the above-laid-out theoretical and scientific foundation. We want to note that this is not intended as a comprehensive analysis of state-of-the-art concepts but rather as a complement to the theoretical and scientific basis with practical insights.

In general, the companies offer their users mobile companion applications (Figure 2.13:i, vii) in combination with (visual) passenger information displays inside the vehicles (Figure 2.13:ii, iii, vi, viii, ix, xii). Auditory displays supplement the visual in-vehicle HMIs with sound signals and voice prompts (e.g., [250]) but also with ambient sounds. For instance, Waymo plays a "relaxing ambient track" [197] when passengers enter the car aiming to provide a comfortable and joyful atmosphere. In line with the general precedence identified by the systematic literature review of Jansen et al. [126], all companies focus on visual and auditory modalities. However, the choice for established technologies generally meets the preferences of prospective AV users derived by Biermann et al. [24]. Similar to conventional (i.e., non-autonomous) mobility-on-demand services, the UIs, and particularly the mobile applications, are designed to accompany users and customers throughout their complete journey – starting from hailing a ride (e.g., 2.13:i) to providing information and support functions during (e.g., 2.13:ii, iii, ix, vii) and after the ride (e.g., 2.13:vi). The claimed goal is to create an "experience that reassures riders every step of the way" [197].

Across the industry, building trust is acknowledged as a critical challenge and a consequent key motivation for system design [8, 197]. To address this challenge, the passenger displays (e.g., Figure 2.13:iii, vi, ix, xii) provide continuous feedback on the

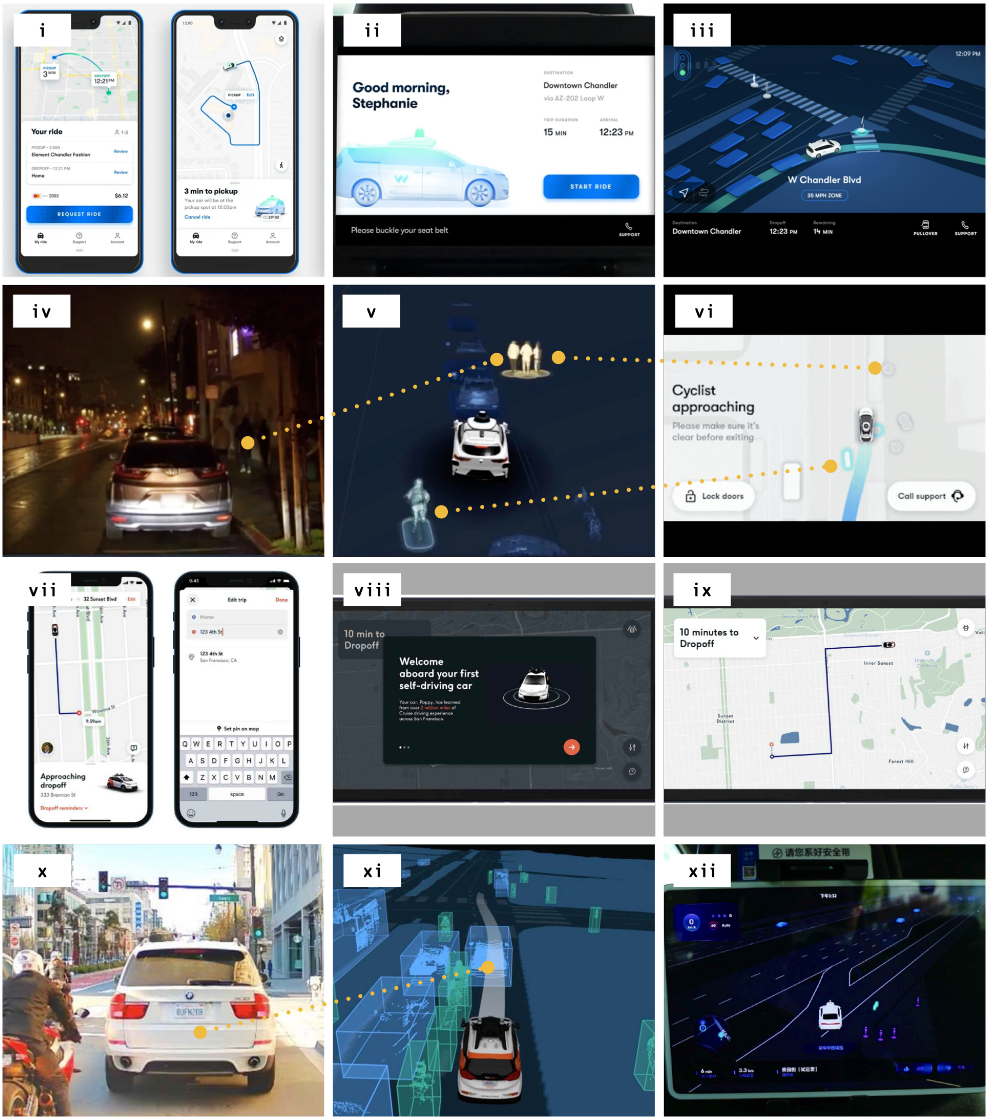


Figure 2.13 – UI examples and scenarios from AV technology companies Waymo (i–vi), Cruise (vii–xi), and Baidu (xii): mobile companion apps (i, vii), passenger displays (ii, iii, vi, viii, ix, xii), and 3D-based visualizations of sensor information used, e.g., for ride monitoring, analysis, and marketing purposes (v, xi). Dotted lines in images iv–vi and x–xi illustrate the connection of objects in the respective ride scenario (see front camera view in images iv and x) to the visualization of the sensor-based information in the UIs. Images i–iii taken from [197], iv–iv taken from [250], vii–ix taken from [53], x–xi taken from [54], and xii taken from [8].

system's status – as recommended by Norman [183] – and combine aHMIs and iHMIs. Besides essential feedback and travel information, all passenger display examples provide map-based route visualizations. The displays shown in Figure 2.13:iii and xii also use 3D-based maps featuring sensor-based visualization of detected objects. In the sense of XAI [105], this information can provide explanations to passengers and tell them what the AV sees and thinks [197]. At the same time, Waymo designer Powel notes that "the passenger doesn't [. . .] want to [. . .] see everything the car sees [. . . but rather . . .] tidy visuals and updates" [197]. This is in line with the discussion from above that the amount of provided information is crucial (see [160, 185]).

Figures 2.13:v and xi illustrate how large amounts of sensor information of an urban ride scene (Figures 2.13:iv and x) can be visualized as a 3D representation. In Figure 2.13:vi, the information on a potentially critical detected object (a cyclist) is further transformed and simplified to a birds-eye-view of the scene, which is complemented with an audio-visual safety-warning telling passengers to take care of the approaching cyclist when leaving the vehicle. Such feedback and warnings might be able to foster passengers' trust in the system, even while not driving. Regarding the 3D representations illustrated in Figures 2.13:v and xi they currently seem to be mainly used by engineers (e.g., for ride monitoring or analysis), safety drivers, or marketing activities. Although their potentially higher complexity, it may also be interesting from both a scientific and a practical perspective to consider such more complex 3D visualizations as an (alternative) UI component for passenger displays to investigate the above-derived question on the right amount (and level of detail) of information and explanations. Also, with regards to (combinations with) other display concepts – e.g., AR-based object visualization (see Chapter 4) – or additional modalities.

The state-of-the-art examples provide a great source of inspiration and interesting potential for further research. Presumably, due to the current implementations of the services and vehicles, the UI examples focus on "private" rides. This may change in the near future due to the great potential of shared rides (Section 2.1.2). In Chapters 3, 4, 5, and 6, we investigate UI concepts for (shared) rides that may provide helpful insights to advance the current state of the art. With the scientific and practical foundation of human-AV interaction design elaborated, we will now have a look at how UI concepts can be evaluated within the HCD process.

2.2.4 Evaluation Methods

Following an HCD approach [124], design solutions such as HMI concepts for human-AV interactions need to be evaluated through the perspective of (end-)users and potentially other stakeholders. Nielsen [178] summarizes four approaches to evaluate user interfaces:

- *Automatic.* Computer-generated assessment of usability metrics.
- *Empirical.* Testing with real users.
- *Formal.* Calculation of usability metrics with models and formulas.
- *Informal.* Using rules of thumb and experienced evaluators.

Since formal methods do not scale well [178] and automatic methods are still limited in their applicability, especially for early development phases, empirical and informal approaches are considerably prevalent. Two widespread approaches to evaluations are inspection-based evaluation and user-based testing [124]. This section provides a concise overview of the two approaches and metrics that are – within the scope of this thesis – suitable for the respective evaluation activities.

Inspection-based Evaluation

Inspection-based methods are a cost-effective way to evaluate design solutions and to identify usability problems [177, 124]. Mostly, inspection-based evaluation uses applicable guidelines (e.g., usability or accessibility guidelines) and requirements to check the design solution against [124]. The methods can vary in the degree of standardization and the type of guidelines used for inspection. Popular inspection methods are heuristic evaluations [177], cognitive walkthroughs, or the inspection regarding the interface's compliance with standards and norms [178].

For instance, heuristic evaluation uses heuristics which can be considered broad rules of thumb [176] as the basis for the assessment. Based on a factor analysis of 249 usability problems, Nielsen derived a set of ten usability heuristics [177] (Table 2.1). Another popular set of guidelines is the seven dialogue principles described in the norm ISO 9241-110 [122], which can be likewise used for inspection-based evaluation (Table 2.1). Those can be regarded as generalized goals for the design and evaluation of HMIs [122]. The ISO guidelines can be considered more abstract, while Nielsen's heuristics are a bit more concrete. Nevertheless, they overlap in their objectives

Table 2.1 – Guidelines for Design and Evaluation of User Interfaces.

Nielsen's usability heuristics [177]	ISO 9241-110 dialogue principles [122]	ISO 15005 dialogue principles [121]
Visibility of system status		
Match between system and the real world	Suitability for the task	Compatibility with driving
User control and freedom	Controllability	Controllability
Consistency and standards		Consistency
	Conformity with user expectations	Conformity with driver expectations
Error prevention		
Recognition rather than recall	Self-descriptiveness	Self-descriptiveness
Flexibility and efficiency of use	Suitability for individualization	
Aesthetic and minimalist design		
Help users recognize, diagnose, and recover from errors	Error tolerance	Error tolerance
Help and documentation		
	Suitability for learning	
		Timing/priorities
		Simplicity

and partially in their formulation. This is generally the case for established design guidelines since they all originate from human psychology and are based on "how people perceive, learn, reason, remember, and convert intentions into action" [129, p. xiii]. More specific to the automotive domain, the norm ISO 15005 [121] offers eight adapted dialogue principles for transport information and control systems (TICS; Table 2.1), which are clustered in the categories 1) appropriate for use while driving, 2) appropriate for the TICS task, and 3) appropriate for the driver. However, as ISO 15005 focuses on driver information and assistance systems, its application for AVs is limited. To our knowledge, there is no adapted norm specifically for (driverless) AVs available yet.

Typically, inspection-based evaluations are conducted by (domain or usability) experts "who base their judgment on prior experience of problems encountered by users and their knowledge of ergonomic guidelines and standards" [124, p. 22]. The inspection can be undertaken by single or multiple evaluators [124, 177]. Having several evaluators assess the design can reduce individual bias [124]. Generally, inspection-based evaluation can be conducted easier and faster than user-based testing, and might find different issues than testings would [124]. It can be complemented with user testing (and, for instance, identify obvious problems beforehand) to make the testing more cost-effective [124].

User-based Testing

Generally, user-based testing refers to the empirical evaluation of design solutions with actual (or prospective) users [177]. It can be conducted at any design stage [125]. At early design stages, artifacts such as scenarios and sketches, or early (e.g., paper-based) prototypes of the design concepts can be presented to users that are "asked to evaluate them in relation to a real context" [124, p. 22]. At later stages, prototypes with higher fidelity are tested by real users to assess, e.g., usability. Therefore, such tests are often referred to as usability tests or – more generally – as user studies. Usually, users are asked to carry out (pre-defined) tasks "using the prototype rather than just be shown demonstrations or a preview of the design" [124, p. 22]. In HCI, evaluation data is often collected by means of observation, measurement, interviews, and questionnaires.

To consider the context of use, user studies can be conducted in the actual environment – which is also referred to as field testing, field study, or field validation [124]. Since the availability of AVs is limited, field studies with actual (i.e., driverless) AVs are hard to realize, especially in the early design stages. However, in both natural and artificial (lab-based) environments, context-based prototyping can be applied to consider physical and social contexts even from early design and development stages and enhance the study's validity (Section 2.3). In the AV domain, this means, e.g. that tests can be conducted in labs with prototyped environments (e.g., with mock-ups, (driving) simulation) or real environments with (simulated) AVs that might use the wizard-of-oz method (Section 2.4).

Often task-based user studies are combined with *thinking-aloud* methods (i.e., users are encouraged to verbalize their thoughts while participating in the test session), (digital or paper-based) questionnaires, and (semi-structured) interviews to assess the current design solution with both quantitative and qualitative methods. Such studies adopt a mixed-method approach [52]. This enables data triangulation: quantitative data (e.g., measured with validated questionnaires, counts of successful task completions, or methods such as eye-tracking), observations, and qualitative data gained from interviews are combined. All studies conducted within this doctoral research (Chapters 3 – 6) adopt such an approach. For the collection of quantitative data, we mostly used standardized questionnaires to assess acceptance (Chen's TAM adaptation [41], Van der Laan et al.'s questionnaire [256]), trust (Körber's Trust in Automation (TiA) questionnaire [141]), UX and usability (User Experience Questionnaire (UEQ) by Laugwitz et al. [150] and its short version UEQ-S by Schrepp et al. [222]), as well as immersion (Igroup Presence Questionnaire (IPQ) [225]) and simulator/motion sickness (Motion Sickness Assessment Questionnaire (MSAQ) [101]). Depending on the study objective, we accompanied the standardized questionnaires and our observations with single-item scales (e.g., on well-being, safety, security, and privacy) and with open and closed questions asked in semi-structured interviews to discuss with study participants their assessment, experiences, and opinions. For valid assessments, context plays an essential role (see [143]). In the following section, we will elaborate on what considering the context in the human-centered design of human-AV interactions means and what potential context-based prototyping holds to create suitable HMI designs and valid (user) studies.

2.3 Context-Based Interface Prototyping

We use the term context-based interface prototyping to describe the approach of prototyping human–machine interactions and respective HMIs in a contextualized setup (see Flohr et al. [85] and Hoggenmüller et al. [118, 117]). In this section, we lay out the theoretical fundamentals and practical considerations regarding context and prototyping in HCI. We also provide an overview of applicable methods for context-based prototyping of (in-vehicle) human-AV interactions.

2.3.1 What is Context?

The notion of context holds a variety of meanings and interpretations, even if we focus solely on the area of computer science and its subdisciplines [218]. Below, we gather existing definitions to render what we consider to be an appropriate understanding for the HCI domain. Starting with a general description, the Cambridge Dictionary defines context in the sense of a "cause of event" as "the situation within which something exists or happens, and that can help explain it" [39]. Schmidt [218] also defines its understanding based on dictionary definitions and uses the term to "describe the environment, situation, state, surroundings, task, and so on" [218, p. 193]. Taking into account several varying definitions from related work, Trivedi and Khanum [253] also derived a rather broad definition and regarded context as "anything which has an effect on the human behaviour" [253, p. 72]. They distinguished cultural, organizational, technological, physical, and social context [253]. Situated within the HCI domain, we focus, similarly to Trivedi and Khanum [253], on the physical and social context and regard the technical aspects as a part of the physical.

2.3.2 Context in Human–Computer Interaction

As can be derived from aforementioned general definitions, context is an essential component in HCI. In ISO 9241-110, the context of use is defined as "users, tasks, equipment (hardware, software and materials), and the physical and social environments in which a product is used" [122, p. 6]. This definition considers users, tasks, and equipment as part of the context which is "surrounded" by physical and social environments. Dey and Abowd [65] channeled their understanding of previous work into an even more tangible definition, which we consider a proper understanding within the scope of this thesis:

> *Context [in HCI] is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.* [65, pp. 3–4]

To consider context in the design and evaluation of systems that are not (yet) feasible or not available, prototypes incorporating this dynamic element can be used. Prototyping can help to understand and explore context and corresponding user experiences, ideas, and concepts [35].

2.3.3 Prototyping as a Means to Consider Context in HMI Design

In line with Thaler's general perspective that a prototype can be "anything that will move the process forward" [247], prototyping can be regarded as "interwoven with nearly all product, service, and systems development efforts" [40, p. 1].

Rendering a methodological, process-oriented perspective, Crabtree [50] distinguished four steps of prototyping: functional selection, construction, evaluation, and iteration. Through evaluation, prototyping enables feedback and communication between the use practice and the design process [50, 158]. For most cases, Crabtree [50] differentiated three interrelated prototyping forms: exploration, experimentation, and evolution. In exploration, prototyping helps to understand context and – although eventually lacking large parts of the desired functionality – helps to foster ″cooperation between designers and end-users“ [50, p. 131]. Experimentation builds on the exploration to demonstrate and refine the prototype pragmatically but still includes user involvement [50, 158]. The evolution phase is focused on the ″development and implementation of a stable prototype [. . . in] the target domain“ [50, p. 131], and thus marks the transformation of the prototype into an actual product situated within its actual context of use.

Lim et al. [154] described prototyping as ″an activity with the purpose of creating a manifestation that, in its simplest form, filters the qualities in which designers are interested, without distorting the understanding of the whole“ [154, p. 4]. This points out that prototyping enables, on the one hand, one to explore (parts of) the final product considering the bigger picture (″the understanding of the whole″), which includes its environmental context. On the other hand, prototyping enables “filtering” to put the focus on particular aspects (“qualities”) of a product, system, or service in which the designer or the team is interested [154]. Filtering dimensions can be, e.g., the prototype's appearance (e.g., in terms of shape, size, and color), considered functionalities, or the degree of interactivity (e.g., in terms of input and output behavior or feedback provision) [154]. Filtering enables the efficient investigation of design ideas without the need to implement the whole thing. Prototype manifestations can, according to Lim et al. [154], differ in three dimensions: material (i.e., the medium used to create the prototype), resolution (i.e., the prototype's fidelity or level of detail), and scope (i.e., the range of what is included in the prototype).

Based on their extensive investigation of various prototyping approaches for urban robotic interfaces, Hoggenmüller concluded that the prototype of a product or system merges with the surrounding (prototyped) context “into one single manifestation” [117, p. 210]. This depicts the interdependence of interface prototypes and their (prototyped) surrounding physical and social environment. Context-based prototyping allows accounting for these circumstances. Besides creating more realistic experiences, it can also reveal requirements and constraints. In the AV domain, this could,

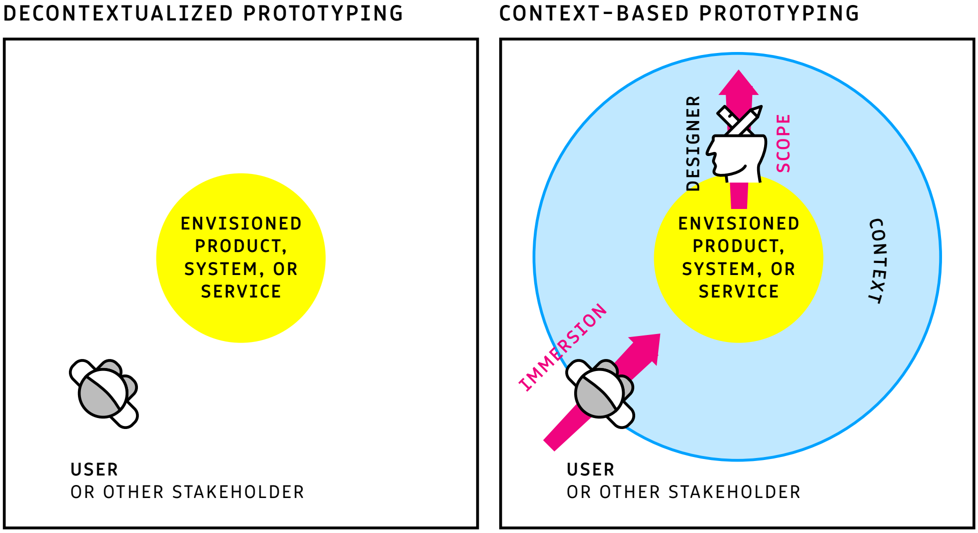


Figure 2.14 – Decontextualized prototyping vs. context-based prototyping. Illustration based on the original diagram of Hoggenmüller [117]. Context-based prototyping situates the envisioned product, system, or service in the context of use. As a result, it increases users' and other stakeholders' immersion and the scope of designers and researchers [117].

e.g., refer to the readability of displayed information, the reachability of controls, or the compatibility of displays and controls. Hoggenmüller [117] illustrated the advantages of context-based (or "contextualized") prototyping – where the envisioned prototype system, product, or service is situated in the (physical or virtual) context – by using a comparison to "decontextualized" prototyping (Figure 2.14). Contextualized prototyping does not only help to increase user's immersion, but also supports designers (and other stakeholders) with envisioning the product, system, or service within the context [117]. With reference to Trivedi and Khanum [253] and Lacher et al. [149], we want to extend this view to also include the social context. Consequently, context-based prototyping can support the HMI design process from all different angles. Before creating (context-based) prototypes, however, a few things should be considered that we will elaborate in the next section.

2.3.4 Considerations and the Impact of Prototyping

The "ultimate goal of prototyping" in the HCI domain is – according to Camburn et al. [40] – the enhancement of performance and UX. Lim et al. [154] postulated as an economic principle of prototyping that "the best prototype is one that, in the simplest and the most efficient way, makes the possibilities and limitations of a design idea visible and measurable." We want to highlight the aspects of simplicity and efficiency as core economic components. Based on the introduced definitions, we suggest expanding this principle to account for the various possible manifestations, places, and users; the different applications of prototypes in the HCD process; and the inherent importance of context. In doing so, we want to note that we do not see prototyping to be a procedure to create something "perfect" or "best." Instead, we consider prototyping more of an iterative tool to achieve specific goals within the overall process. As such, it can provide a certain amount of (maximum) value but without achieving something like a state of perfection. In line with this, Camburn et al. [40] also pointed out that iterative prototyping directly causes an increase in performance and the ability to meet (difficult) requirements. The following statement summarizes these contemplations as our working definition. References to Lim et al. [154] and ISO 9241 [124] are highlighted in italics.

> Prototyping in HCD is of greatest value when it *most simply and efficiently* supports achieving the goals of a particular activity, such as *analysis, design, or evaluation*. Prototyping can, for example, make ideas, concepts, and contexts *visible*, tangible, or *measurable*.

To achieve this, there is a wide variety of methods, materials, and tools available (Section 2.4). However, when it comes to their selection, it is important to consider that "everything is best for something and worst for something else" [38]. We render this to be crucial for successful (context-based) prototyping and agree with Buxton that the "trick is knowing what is what, for what, when, for whom, where, and most importantly, why" [38]. Similar to that, Dodge [70] lays out that how much a prototype can teach us depends on what, why, and how we are prototyping; and the when (i.e., the point of time in the process) and the amount of time spent to create the prototype significantly affect the impact on the (final) design. What, why, and how can directly refer to the notions of manifestations and filters introduced by Lim et al. [154]. Dodge [70] formalized their relationship with the time spent and the point of time in the process (when) in Equation (2.1).

$$\frac{\text{What} \times \text{Why} \times \text{How}}{\text{Time spent}} \times \text{When} \underset{Dodge}{=} \text{Impact on design} \tag{2.1}$$

We have already elaborated on the materials and the scope of prototypes – i.e., two of the three introduced dimensions of prototype manifestations described by Lim et al. [154]. Now, we want to complement this with a view on the third dimension, the resolution, or fidelity. In the HCI domain, fidelity is often referred to as "the extent to which a computer application or system reproduces visual appearance, interaction style and functionalities" [58, 208]. In other words, fidelity describes the "level of detail" [58] or the degree to which a prototype represents the (actual, planned, or final) product, system, or service. In the above equation, fidelity is addressed by what, how, and where. Often, prototypes are described with the dichotomous categories "low-fidelity" or "high-fidelity" [208]. However, as Virzi et al. [259] and Warfel [262] pointed out, fidelity should be regarded as a continuum, not as a dichotomy. The required fidelity depends on the goal or the purpose that one wants to accomplish with a prototype [262]. Basically, this means that it depends on the answers to the questions that are part of Doge's equation [70] and on the (requirements of the) target audience. With regard to the use of prototypes as (part of) simulations, Dahl et al. [58] distinguished three components that contribute to overall simulation fidelity: prototype fidelity, environment fidelity, and psychological fidelity. While we value this differentiation, we want to note – considering our adopted working definition of prototyping – that an environmental (or contextual) representation as part of a simulation can also be considered a form of a prototype. In general, higher fidelity often results in higher efforts to produce a prototype. Depending on the objective of, e.g., a study, it is vital to select an appropriate level of fidelity – since it can affect the accuracy of others' interpretations [40] and may affect participants' immersion, and consequently, their assessment [117] – and of course, economic aspects.

From an HCI perspective, AVs and related mobility (on-demand) concepts are still in an early development phase (when). Especially in this early stage, context-based prototyping can have a substantial impact (Equation (7.1). Section 2.4 provides a qualitative semi-systematic literature review [236] and discusses suitable methods to inform future research on what, why, how, and where context-based prototypes of AV HMIs can be created.

2.4 Methods for Context-Based Prototyping of In-Vehicle Interactions

Testing and evaluating new AV HMIs and concepts in early development phases with actual AVs and in real traffic is – similarly to the development of advanced driver assistance systems [33] – only possible with tight limitations. Aside from current technological constraints, this is primarily due to ethical aspects, especially regarding the potential danger when involving participants, other road users, and infrastructure. Context-based prototyping can help to face the problem of AVs being still in their technical infancy.

This section provides an overview of applicable methods for context-based interface prototyping [85, 118, 117] of in-vehicle interactions with AVs. Such methods enable researchers, designers, and other stakeholders to establish and experience contextualized setups of human-AV interactions to consider environmental factors in HCD activities, such as analysis, design, and evaluation. We focus on methods that support prototyping interfaces within their (intended) physical and social context. However, we do not focus on concrete interface prototypes. Within the scope of this thesis, we neither discuss differences between HMI prototypes, e.g., in terms of the fidelity of sketches, wireframes, and high-fidelity visual design prototypes. Nor do we look at prototyping tools such as Sketch, Figma, Antetype, or Adobe XD. Interface prototypes are, especially with regard to the above-cited definitions, of course, a crucial part of the context of use. A detailed discussion of these is, however, beyond the scope of this dissertation.

Depending on the contextual situation, some methods are more suitable than others – for instance, regarding economic aspects. For example, for some AV scenarios, it might be required to have a dynamic high-fidelity representation of a ride through an urban environment. For others, it can be sufficient to have a static mock-up of a vehicle environment in a laboratory – e.g., to evaluate the general placement of display concepts within a vehicle. Often, a combination of several methods is used – e.g., Flohr et al. [85, 86] used immersive video to create a dynamic ride simulation in combination with both interactive and video-based interface prototypes and actors that simulated the social context.

In general, prototyping methods need to be assessed and chosen depending on the aim or purpose of a particular project while considering their limitations. For example, while

VR setups offer high degrees of flexibility, it still needs to be determined how participants' perceptions differ from reality [96]. With reference to Bengler [17], Fuest [96] proposed that each method needs to be assessed with the three scientific quality criteria:

- *Objectivity.* The extent to which results are independent of any influences outside the matter of subject [198]. E.g., independence from influences of the investigator [96], test moderator, or analyzing and interpreting person [198].
- *Reliability.* The accuracy with which something is measured [198] or executed. E.g., for a wizard-of-oz study, the same driving style needs to be reproduced for each session [175, 96].
- *Validity.* The extent to which a method actually measures or predicts what it is supposed to [198]. In HCI, one often distinguishes between internal validity – i.e., the extent of control of a study or method [235] – and external validity, which often refers to the generalizability of results [235].

Concerning the (external) validity of prototyping methods, and in particular, simulation methods, essential aspects are study participants' immersion and sense of presence in the context, i.e., in the simulated environment or virtual world. A virtual world can be described as "an imaginary space often manifested through a medium" ([231], p. 8). The "sensation of being in an environment" ([231], p. 10) such as this virtual world is described by the term immersion. Creating immersion and the related experience of presence is a significant challenge in simulator studies [33] or – more general – in context-based prototyping. Sherman and Craig differentiated mental immersion as a "state of being deeply engaged" from physical immersion as "bodily entering into a medium [and the] synthetic stimulus of the body's senses via the use of technology" ([231], p. 10). However, this "does not imply [that] all senses or that the entire body is immersed/engulfed" ([231], p. 10) at the same time. While sense of presence is often used as a synonym to immersion, Sherman and Craig assigned presence as equivalent to the state of mental immersion [231]. Similarly, Bubb [33] described presence as a cognitive state where users have the impression of being part of the virtual world which can be achieved through suitable technology design.

With these criteria introduced, the following sections provide an overview of the most frequently used context-based prototyping methods for human-AV interaction. Besides related work, we also included the setups and methods, we used in Chapters 3, 5, and 6. We focus on the application for empirical studies on in-vehicle HMIs (and do not consider, e.g., questionnaire-based online studies or external HMIs – though some of the methods may also be applied for such study designs), but still, note that this collection is not exhaustive. We cluster the methods into the categories 1) static mock-up (including spatial interior and exterior representations), 2) ride simulation (with a focus on virtual and mixed reality and (immersive) video), 3) social context simulation, 4) wizard-of-oz, and 5) experimental vehicle. Table 2.2 summarizes the methods' key aspects and practical challenges.

Table 2.2 – Context-based prototyping methods for human-AV interactions and their challenges.

Method	Description	Challenges
Static mock-up	Static elements – that do not show (dynamic) changes over time – provide a spatial representation of AV interior and exterior components.	Weighing (study) requirements (e.g., regarding fidelity) and effort to construct a mock-up.
Ride simulation	Simulating the dynamic (physical) context of riding in an AV using VR, MR, (spatial) sound, (immersive) video, or a combination of the mentioned approaches.	Achieving a sufficient level of fidelity for participants' immersion in the simulation [33] that is required for a valid contextual representation; Coping with occurring simulator sickness symptoms [5, 116].
Social context simulation	Simulating interactions and communication with others, e.g., co-passengers in shared AV rides. Prototypes can, e.g., make use of sound and actors [85] or enactment and props [228].	Incorporating social context into (semi-)controlled test environments can lead to adverse effects. E.g., people might feel uncomfortable with other (unknown) people present in certain situations, which might also lead to adverse effects in terms of simulator sickness [85].
Wizard-of-Oz (WoOz)	Making participants believe they are riding in a real AV while a human driving wizard controls the vehicle [19]. WoOz can be used to prototype AVs on test tracks and on public roads.	Keeping up the WoOz deception throughout the conduct of the study; Coping with varying environmental conditions (e.g., weather, other road users) and ensuring comparability of test rides [175, 19].
Experimental vehicle	Vehicles with actual (but to some extent limited) automated driving capabilities [96].	Counteracting (technological) limitations that might not meet participants' expectations (see, e.g., [181]) and affect their assessment, e.g., limited speed, restricted areas, presence of a safety driver.

2.4.1 Static Mock-Ups

We use the term static mock-up to categorize methods that enable the inclusion of static contextual elements that do not show any changes over time, e.g., in their appearance. In terms of in-vehicle human-AV interaction, static components can, for instance, be used to spatially prototype the interior and the exterior of an AV. The construction of a mock-up can directly affect participants' sense of presence [33].

To analyze user requirements regarding the design of shared AVs, Schuß et al. [228] used a tent-based mock-up to create an enterable prototype of an AV's exterior. This also enabled them to situate participants of an empirical user study in a closed environment resembling the "pod'-like interior of a shared AV (Figure 2.15). Conventional chairs were used to resemble the seats of the AV (Figure 2.15ii). Similarly to that, we used office chairs to do the same as part of an immersive video-based setup and used room walls, a movable whiteboard, and wooden pallets to create a rudimentary spatial mock-up (Figure 2.15iii,v; Chapters 3 [85], 5 [86], and 6 [87]).

While the mock-up of Schuß et al. [228] was constructed using metal poles and a tent canvas, other materials such as paper, cardboard, and image prints could also be used to (re-)create similar setups. Static setups without additional components such as dynamic simulation, including the ones by Schuß et al. [228] and the ones we used in Chapters 3, 4 and 6, can be considered on the lower side of the fidelity continuum. However, exterior and interior prototypes also allow approaches with higher fidelity to provide a basis for the creation of sophisticated prototypes with the aim of resembling the final vehicle design (e.g., Figure 2.15v). Items, props, and physical controls such as emergency stop buttons or breaks can extend the physical immersion. They also enable one to investigate the match between digital displays and respective physical controls and to evaluate their compatibility or corresponding constraints. Furthermore, especially personal items can also increase social context simulation (Section 2.4.3).

2.4.2 Ride Simulation

Simulators enable researchers and designers to consider dynamic contextual factors in early development stages. In terms of driving and ride simulation, such dynamic factors might compromise, e.g., seeing a changing urban environment while looking out of a vehicle window during a ride through a city, sounds of the vehicle when accelerating, or the behavior of other road users, such as other vehicles or pedestrians. Depending

Figure 2.15 – Examples of static interior mock-ups of (shared) AVs in combination with other prototyping methods. Images (i, ii) were taken from Schuß et al. [228] (CC BY 4.0), who used a tent-based setup in combination with a dynamic enactment simulation and physical items. Images (iii–v) show video-based simulation setups with different spatial mock-up components (chairs, walls, tent, whiteboard, wooden pellets) we used for the setups described in Chapters 3 [85], 5 [86], and 6 [87]. Image (vi) shows a high-fidelity interior of a shared AV.

on the setup and research aim, simulators offer, on the one hand, high controllability (e.g., of environmental conditions) and reproducibility (e.g., created simulations and test parameters can be easily transferred to other studies) [219, 61]. On the other hand, they provide high flexibility in their manifestation and in terms of simplicity in data collection [219, 61]. Furthermore, they allow the safe assessment of new systems and interfaces in early development phases [61]. With regard to the aforementioned quality criteria, simulators provide an excellent basis for the reliability of a study.

However, a major challenge of using simulators is the creation of a participant's experience of presence in the simulated environment [33]. To achieve high presence perception, high-fidelity reproduction of visual, acoustic, haptic, and spatial stimuli is required [33]. As limitations regarding the realistic representation of these stimuli persist even in modern simulators, the validity of (automated driving) simulator studies remains an important research topic [116]. Furthermore, so called simulator-sickness symptoms, such as nausea, vertigo, sweating, or headaches [116], might occur while being in a simulated environment. Almallah et al. [5] found that women are more prone to simulator sickness than men and that older people experience more severe symptoms. Simulator sickness was found to be related to the sense of presence [5]. I.e., the more people are immersed in a simulation, the less likely is the occurrence of simulator sickness. According to Bubb [33], immersion depends on the reproduction quality of spatial and temporal stimuli that humans perceive with their sensory organs. Almallah et al. [5] also found that urban environments with close buildings and lower speed limits can increase participants' sense of presence while simultaneously decreasing the occurrence of simulator-sickness symptoms. Hock et al. [116] provided a checklist to overcome typical challenges when conducting (driving) simulator studies.

As mentioned before, we focus on prototyping the physical and social context of in-vehicle human-AV interactions. By simulating these contextual aspects, we consider visual and auditory (noise, sound) impressions to be most relevant for (cost-effective) context-based prototyping, and will therefore put an emphasis on these. However, we want to note that other components, such as motion simulation and the inclusion of vehicle dynamics, might also be vital for some scenarios. Most common simulators used in automotive HCI research immerse study participants in a virtual world by using either computer-generated imagery (CGI; e.g., [100, 82, 271]) or (immersive) video (e.g., [143, 145, 98, 85, 86]). Below, we discuss these approaches, their theoretical background, and their application in current HCI research.

Virtual and Mixed Reality

Being immersed in an alternate reality such as a virtual world is usually referred to as virtual reality (VR) [231]. VR allows investigations about how humans interact with computer-created worlds and simulations [12]. Milgram and Kishino describe a VR environment as "one in which the participant-observer is totally immersed in, and able to interact with, a completely synthetic world" [168, p. 2]. To describe how VR and associated concepts are related, they introduced a continuum between the real environment and the virtual environment, where they describe the space in between as mixed reality (MR) (Figure 2.16). MR is regarded as a state in which "real world and virtual world objects are presented together" [168, p. 2]. Subsets of mixed reality are augmented reality (AR; in which the real environment is supplemented with virtual (computer-generated) objects [7]) and augmented virtuality ("in which real objects are added to virtual ones [. . .] and the surrounding environment is virtual" [7, p. 34].

Figure 2.16 – Milgram's and Kishino's reality-virtuality continuum. Illustration based on the original diagram of Milgram and Kishino [168]. Prototypes can make use of the whole spectrum to make products, systems, and services experiential.

As can be seen in related work, e.g., in Azuma et al. [7], VR is often used as a synonym for virtual environments. Generally, both VR and AR, and augmented virtuality, can be used for simulation and context-based prototyping. For instance, Morozova [173] presented a "mixedUX" prototyping framework for usability testing in AR. While augmented virtuality is quite rarely used in the AV domain, some studies use AR to investigate new HMI concepts. Haeuslschmid et al. [111], for example, used AR to prototype interactions with a virtual avatar. However, most state-of-the-art driving/ride simulators use CGI-based VR as their basis.

In automotive simulators, popular hardware setups are CAVE-like [55] environments, head-mounted displays (e.g., HTC Vive, Oculus Rift, and Microsoft HoloLens), or compilations of three monitors (Figure 2.17). Currently, these methods are often applied to evaluate systems such as advanced driver assistance systems for non-driverless vehicles (i.e., SAE levels 0 – 3) or for teleoperation of vehicles in combination with video live-streams of their environment (e.g., [72]). Simulator studies in the automotive domain mostly refer to VR setups created with CGI (computer-generated imagery). However, an immersive virtual environment can also be created using real-world videos [143, 98, 85], into which we will take a more in-depth look in the following section.

Video and Immersive Video

Instead of CGI-based VR, it is also viable to use real-world videos as a basis for the simulation. For instance, Krome et al. [145] and Haeuslschmid et al. [111] used single videos from real-world traffic situations to provide a basic representation of the physical context of a ride through an urban environment for their HMI studies. Real-world footage can be enhanced with additional imagery or CGI – e.g., to prototype AR-based avatar concepts [111]. Multiple real-world videos can also be used to

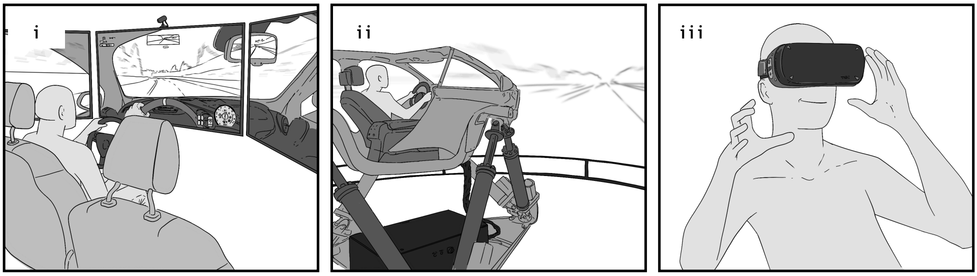


Figure 2.17 – Typical simulator hardware used in the automotive domain: (i) compilation of three displays and an interior mock-up, (ii) vehicle (mock-up) in a CAVE-like environment – here, with hexapod-based motion platform, (iii) head-mounted VR display. Illustrations provided by our colleague Matthias Rebmann.

create a more immersive spatial simulation, similar to a CAVE-like [55] environment (see [143, 186]). Kray et al. [143] called this approach immersive video. It features a high-fidelity audio-visual representation of real-world contexts and a high degree of control. Aside from using multiple cameras, it is also viable to use special equipment, such as 360° cameras.

Gerber, Schroeter, and Vehns [98] constructed, for instance, a 3D-printed camera rig to align three action cameras to capture 180° video footage. Since the CGI-based virtual environment of their pre-existing dynamic driving simulator with three front screens and a field of view of 180° lacked the required level of contextual fidelity and detail, they chose to use immersive video instead of CGI for their automated driving studies (SAE level 2 and 3) [98]. While their setup's overall immersion was assessed to be high, Gerber et al. [98] stated that the sense of immersion was different compared to CGI-based studies, but that participants' familiarity with the local environment supported their feedback quality.

In Chapter 3 [85], we present a straightforward and low-budget adaption of an immersive video approach for the AV domain (SAE levels 4 and 5) based on the works of Kray et al. [143] and Gerber et al. [98], but without the need for special equipment such as camera rigs or sophisticated simulator setups (see also Figure 2.15iii).

2.4.3 Social Context Simulation

Simulating the social context means prototyping (aspects of) the context of interacting and communicating with other people. In shared AVs, and more generally, in public transportation, passengers face encounters with other people, such as co-passengers. Inevitably, they communicate with each other – even when they do not intend to do so since "one cannot not communicate" [263, p. 30]. This means that even though humans do not communicate verbally with others, they still communicate implicitly, e.g., through behavior, gestures, or mimics. Since people's trust in a system does also depend on (their trust in) other people [149], and the presence of co-passengers can affect people's wellbeing [85] and perceived security [228], considering the social context is a vital aspect for valid context-based prototyping.

Actors and Enactment

To derive design implications for human-AV interaction in an interview study, Schuß et al. [228] embodied participants with a user enactment [184] part in the user journey of using shared AVs for transportation. Odom et al. described user enactment as "fieldwork of the future" and as a method where "designers construct both the physical form and the social context of simulated futures, and ask users to enact loosely scripted scenarios involving situations they are familiar with, and novel technical interventions designed to address these situations" [184, p. 338]. First, Schuß et al. [228] told participants to imagine typical scenarios where they would ride with a shared AV in the future (e.g., traveling home from work, picking up kids from school). Then, they asked participants to enact and to interact with a static mock-up as if they would conduct a ride with a shared AV. In this shared ride, an actor joined the ride and mimicked a stranger with whom participants shared their ride [228]. The authors concluded that the enactment supported the consideration of the context of use [228]. Similarly to that, we prototyped the social context of a shared ride, i.e., the interaction with fellow passengers, by using an actor who simulated another passenger getting on and off a shared ride. The results presented in Chapter 3 [85] suggest that actors can increase participants' immersion in the simulation but can also affect their wellbeing. Other approaches for social context simulation might encompass, e.g., the use of mannequins, puppets, drawings, or AR overlays to simulate the physical presence of others in a real environment.

Items and Props

Additional physical items can enhance the simulation of social context. For instance, Schuß et al. [228] let participants choose fitting props and items (e.g., backpacks, books, laptops, a stroller, or a baby doll representing a child) to take with them along the enacted journey in the shared AV (Figure 2.15) with the aim of enhancing the realism of the simulation and immersion. Besides using personal items for social environment simulation, physical items might also concern assistive devices such as wheelchairs or glasses that may be relevant for accessibility-related design and research activities.

Sound

Besides the visual environment, auditory aspects play an important role in terms of both the physical and the social context. This can encompass, e.g., driving noises, noises from other vehicles, or sounds of co-passengers. In an (immersive) video-based simulation, the sound footage of a driving video could serve as a reasonable basis. However, this might be extended meaningfully with additional sounds to simulate specific scenarios, including aspects of the social context, such as noises of passengers getting on and off a shared AV, or sounds of opening and closing vehicle doors (Chapters 3, 5, and 6). When used without visual (VR) simulation, sound simulations can either stand-alone or be used in combination with mock-ups. In such cases, these can be regarded as a form of AR considering the above-discussed works of [7, 168, 231].

2.4.4 Wizard-of-Oz

Wizard-of-Oz (WoOz) studies allow for the evaluation of intelligent systems such as AVs prior to their availability [19]. They can go beyond the limitations of laboratory or test environments [261]. The general idea behind the WoOz method is to make participants believe that they are interacting with an intelligent artificial system. At the same time, their internal workings are, in fact, simulated by humans – the so-called wizards [23]. When using the method to prototype AVs, study participants are led to believe that a vehicle is driving (fully) automated while they are actually driven by a human driving wizard who controls it [19]. WoOz can be used to prototype AVs and corresponding interfaces in real-world environments, i.e., on public roads [64, 131, 135, 167, 261].

Figure 2.18 – Example of a video-based Wizard-of-Oz setup that we created for our in-vehicle interaction studies (Chapter 4) inspired by the works of Karjanto et al. [131] and Detjen et al. [64].

The past decade saw a significant increase in the popularity of WoOz within the automotive domain, e.g., to evaluate new HMI concepts [113] or to investigate non-driving related activities [64]. As a result, Bengler et al. [19] proclaimed the "renaissance" of WoOz [19] and provided an overview of typical WoOz settings. Those vary depending on the automation level of the envisioned system and the degree of participants' (illusion of) control. Given that in AVs, passengers are only passive occupants, "classic" controls such as steering wheels and pedals are not required for (most) AV studies. Common setups typically position participants on the co-driver's seat in the front [11, 135, 261] or in the back [64, 131, 167]) while physically separating them from the driving and interaction wizards. Karjanto et al. [131] and Detjen et al. [64], for instance, positioned study participants in the back of their WoOz vehicle and used an isolator wall with a mounted TV displaying the video stream of a webcam installed on the vehicle's windshield. Inspired by their setups, we also created a video-based WoOz setup for our in-vehicle interaction studies (Figure 2.18; Chapter 4 [88]).

While offering the advantage of relatively low limitations in terms of the physical context, WoOz poses methodological challenges. Concerning a study's validity, it is essential to guide participants to believe in the WoOz illusion and to have the vehicle behave like an actual AV would do [175, 261]. To achieve this, human driving wizards need to follow a pre-defined driving style strictly (e.g., like "a professional limo driver" [11, p. 285]). This style must be consistently reproduced by the wizard(s) throughout all sessions and test rides to support the reliability of the study [175].

Cover stories [64] are used to create and maintain the WoOz illusion. In such cover stories, participants are told about the (simulated) capabilities of the AV, e.g., driving autonomously in an urban environment. Varying environmental conditions such as traffic density, the presence and behavior of other road users, weather, and lighting conditions, poses further challenges in terms of reliability [175]. Such variations might impact study goals and the comparability of test rides [19]. Bengler et al. [19], thus, proposed to include an assessment of the "comparability of test drives and the believability of the illusion" when conducting WoOz studies.

2.4.5 Experimental Vehicle

As mentioned before, "actual" AVs are still under development and currently only available with limitations. We use the term experimental vehicle [96] to describe vehicles with actual automated driving capabilities. Such experimental vehicles typically come with limitations, such as (maximum) speed limits, restrictions to specific test scenarios and tracks, and/or the need for constant surveillance through a human safety driver. Like WoOz vehicles, experimental vehicles can be used both on test tracks and in real traffic [96].

As this approach requires an actual vehicle and the technical expertise for implementing the desired scenarios, it is expensive – particularly compared to other prototyping methods [96]. Apart from the high cost, and since participants experience an actual ride in an automated vehicle that – potentially – takes place on a real (public) road, the method promises to offer high validity. Likewise to WoOz, reliability is impaired due to the dynamic and uncontrolled environment [96], e.g., regarding real traffic and weather. It needs to be considered that the limitations might also affect study participants' evaluation. For instance, they might assess certain aspects, such as their trust in the system and safety perception, differently with the knowledge that the study is conducted in a restricted area or that there is a safety driver present. Furthermore, as, e.g., Nordhoff et al. [182] pointed out based on their results of an interview study with 30 participants experiencing a ride in an experimental vehicle on a campus in Berlin-Schöneberg: the experimental vehicle might not meet participants' expectations – which might then again affect their assessment.

2.5 Chapter Summary

This chapter reviewed the background and related work on human-AV interaction and context-based interface prototyping. We started with an outline of AVs' technical fundamentals and challenges and looked at plausible implementation scenarios of private and shared AVs in our daily life. With this background, we focused on human-AV interactions. Here, we particularly investigated acceptance and human factors challenges that need to be overcome for the technology's adoption. In this matter, trust in AI-based technology was identified as one of the most critical adoption barriers [41, 132]. Besides trust, we also saw that acceptance-related factors such as ease of use, usefulness, and enjoyment impact people's intention to use AVs. Corresponding factors regarding pragmatic and hedonic UX, situation awareness, safety/security concerns, and privacy issues play further relevant roles.

As a foundation for developing suitable human-AV interactions and corresponding HMIs, we argued for applying the HCD process described in ISO 9241-210 [125]. After getting an overview of HMI types for internal and external communication, passengers' information demands, and interaction concepts, we discussed HCAI and XAI as suitable frameworks to discuss and account for prospective AV users' needs. Particularly, we identified transparent system feedback as a crucial aspect for addressing concerns related to trust and security. In terms of **RQ1**, we should generally aim for an "experience that reassures riders every step of the way" [197].

To counteract the mentioned hurdles from a methodological point, we identified context as a crucial aspect to consider in all HCD activities. Consequently, we discussed the application of context-based interface prototyping [85, 118, 117], which concerns the creation of contextualized setups that enable designers, researchers, and other stakeholders to consider environmental factors through various activities, including analysis, design, and evaluation. Regarding **RQ2**, we argue that context-based prototyping is required to consider the complex interrelations of human-AV interactions with influencing factors from the environment – specifically in the early stages of development. Particularly for evaluation activities, we pinpointed inspection-based and user-based assessments as the most widespread and suitable methods.

Before diving deeper into context-based prototyping for human-AV interactions, we first derived a precise understanding of context and prototyping based on previous definitions and perspectives from fundamental literature and related work. For the notion of *context*, we adopted the definition of Dey and Abowd [65], who define context in the field of HCI as any information that characterizes the situation of a person, place, or object relevant to the interaction. We saw that context can encompass physical factors but also social, cultural, or organizational factors [253]. Therefore, within the scope of this doctoral research, we focused on the physical and the social context since we deemed them to be the most critical facets for prototyping HMIs for (shared) AVs. With regards to *prototyping*, we discussed the framework of Lim et al. [154], who regard two dimensions of prototyping: prototypes as manifestations (e.g., of design ideas) and prototypes as a method to filter investigated aspects (e.g., particular functions of a concept) [154]. In contextualized prototyping, the prototype of an interface, system, or service merges with the context "into one single manifestation" [117, p. 210]. Based on the work of Lim et al. [154] and the ISO 9241 [125], we derived that prototyping offers the most value when it supports achieving the goals of a particular human-centered design activity simply and efficiently. To achieve this, researchers and practitioners should consider what, why, and when prototypes are required to decide how (and with which fidelity and resources) a specific prototype should be manifested.

We then transferred this theoretical understanding to the prototyping of human-AV interactions. Since AVs do not require the presence of a human operator or driver, passengers will solely interact with the autonomous system and other passengers. This means, they do not (need to be able to) directly control the vehicle, which is an essential prototyping consideration in contrast to lower automation levels and enables straightforward prototyping methods. We rendered the scientific quality criteria objectivity, reliability, and validity to assess methods and conducted research. Finally, the chapter concluded with an overview of suitable methods for context-based prototyping of human-AV interactions: 1) static mock-up, 2) ride simulation (including virtual and mixed reality as well as video-based approaches), 3) social context simulation (including actors and enactment strategies, besides social items and props), 4) wizard-of-oz vehicle (in which a human driver simulates an AV on a test track or in real traffic), and 5) experimental vehicle (i.e., an actual (automated) vehicle with restrictions).

3

Immersive Video-Based Simulation of Shared AV Rides

To defy AV acceptance hurdles, a precise understanding of the interrelation between people, system, and environment is needed [149] while human factors and user requirements need to be considered from early development phases on [31]. Suitable analysis, design and evaluation methods, and particularly appropriate prototyping approaches, are required to inform and enable both, researchers and designers. Contributing to the development of such approaches and our general research question RQ1, this chapter explores prototyping and evaluation methods as well as human factor challenges with a focus on physical and social contextual simulation and HMIs for (S)AVs, e.g., concerning passenger information systems in a public AMoD system.

THE CONTENTS OF THIS CHAPTER WERE PUBLISHED PREVIOUSLY AS PART OF:

[85] **Lukas A. Flohr**, Dominik Janetzko, Dieter P. Wallach, Sebastian C. Scholz, and Antonio Krüger. 2020. Context-Based Interface Prototyping and Evaluation for (Shared) Autonomous Vehicles Using a Lightweight Immersive Video-Based Simulator. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1379–1390. `https://doi.org/10.1145/3357236.3395468`

First, we introduce and evaluate a simple method for creating a low-budget (S)AV simulator. In contrast to common driving simulators, our approach does neither have any controls related to the driving-task nor rely on a ‘pure’ virtual environment in terms of CGI. Instead, real-world video and audio footage is used to create an audiovisual simulation. Second, we investigate social context simulation, particularly the influence of an actor – imitating an entering and leaving passenger – on presence perception, simulator sickness, and participants’ assessment of prototypes and systems.

3.1 Video-Based (Shared) AV Simulator

We consider – similar to Krome et al. [145] – investigating the future context of driving in a (S)AV as a ‘prototyping challenge’. To solve this challenge, we created an immersive video-based AV simulator (Figure 3.1). The simulator consists of a CAVE-like environment that was created by setting up three video projectors, a stereo sound system and a 3×2 seating group in an office room. The video footage was recorded using three action cameras while driving through urban traffic and postprocessed to create a synchronized immersive video. Using immersive video holds two major advantages: it provides a high-fidelity representation of the real world and the creation of the simulation does not require programming skills. Basically, our setup makes riding in a driverless pod-like AV experienceable and provides the basis for context-based user research, interface prototyping and usability testing. Interfaces – e.g., passenger information displays – can be evaluated in a controlled environment including high-level contextual information. In contrast to the setup by Gerber, Schroeter, and Vehns [98], our approach focuses explicitly on simulating a (shared) AV (SAE levels 4 and 5). However, it can be regarded as a modified adaption of their ’IVAD’ simulator [98]. By using relatively low-budget consumer equipment only, we place particular emphasis on simplicity, reproducibility and cost efficiency.

3.1.1 Setting up the Simulator

As AVs are driverless, it is not necessary to have control elements like a steering wheel or a gas pedal. Thus, the simulator interior can be rather simple and abstract. We – for example – use a standard office room as a basis to make the setup easy to reproduce

Figure 3.1 – Immersive video-based AV simulator with three video projections and a 26-inch passenger information display. In the conducted user study, an actor (left) joined the simulated ride of the participant (right) to simulate the social context of shared AV rides.

in any kind of typical (office) building. Our sight simulation encompasses a viewing area of about 270° (Figure 3.2, 3.1) and is displayed by three video projections (Vivitek DH833 with 1080p resolution). The projections resemble the view out of the front window (projection size: 96.5 inch × 41.34 inch) and the side windows (projection size: 72.8 inch × 41.34 inch) of a "pod"-like people mover, e.g., [75]. We also considered using large monitor displays instead of video projectors for the wall-size simulation but decided to go with the latter because they are both, less obtrusive while not in use and less expensive.

A stereo sound system (Fostex PM0.3d) displaying the acoustic simulation and sound signals accompanies the visual simulation. For testing purposes, the setup can be extended by HMI displays and controls as well as by seating groups and other components resembling the interior of the respective vehicle. In our case, we want the simulator to look like the interior of a people mover used for public transport with a 3×2 seating group in the front area. Figure 3.2 schematically illustrates our final setup.

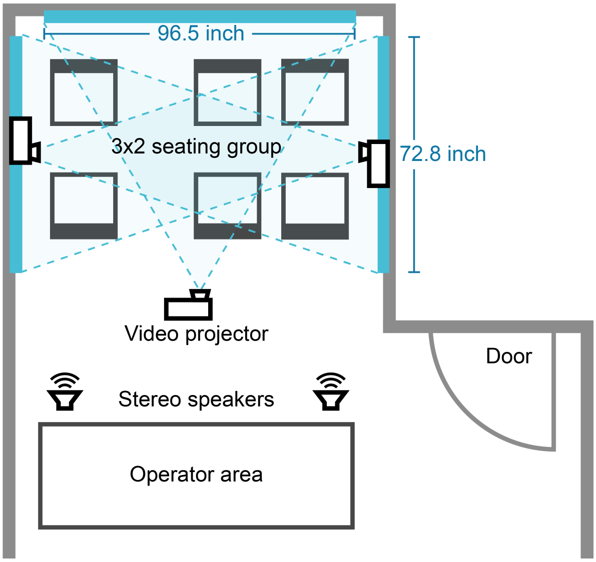


Figure 3.2 – Schematic illustration of the AV simulator setup consisting of three 1080p video projectors, stereo speakers, and a 3×2 seating group.

3.1.2 Recording Video and Audio

The video and audio footage necessary for the basic simulation can be recorded anywhere by driving a common camera-equipped car through traffic. There is no need for special or expensive equipment. However, the route (and potential stops) should be carefully planned in advance to meet the requirements of the intended (prototyping or evaluation) activities. In case of simulating a shared ride, the AV needs to stop sometimes to enable other (simulated) passengers to enter or leave the vehicle. It is recommended to devise a plan for adequate spots for such breaks in advance and to determine parameters like stopping time before starting the ride. To mimic the behavior of an AV, the driving style should be very conservative, highly anticipatory and conform with the traffic rules [98]. Additional recordings of GNSS data, e.g., with a smartphone navigation app, help to synchronize HMI information afterward.

For the video recordings, three identical low-budget action cameras (Crosstour CT8500, 4K resolution, 170° wide-angle fisheye lens) were used. The cameras were mounted with suction cups on the car's windscreen, left window and right window (Figure 3.3). To later create an immersive video out of the recordings, the cameras need to be aligned to each other regarding position, height, orientation and color

Figure 3.3 – Recording footage of urban traffic for the audiovisual simulation using low-budget action cameras.

configuration. No special equipment, e.g., a video rig [98], was used, in order to keep the setup effort as minimal as possible. In our case it worked best to simply position the cameras on about the same height and at the center of the three windows (Figure 3.3). Then, they were manually fine-calibrated in terms of alignment and orientation with a small overlap area at the edge of the videos, while monitoring their streams on a 10.1-inch tablet.

Certain challenges arise when placing the cameras behind the car's windows, e.g., reflections, occlusions and rain drops or dirt on the windows might impair video quality. To minimize disturbing artifacts, some preparations are necessary. Reflective interior elements should be covered with non-reflective materials, e.g., dark tape. Drivers should wear long dark clothes to diminish reflections of body parts. Windows and camera lenses should be kept clean at all times. To avoid extreme lighting, raindrops or other adverse effects, the recording time should be carefully chosen with regard to the weather conditions. In support of [98], recording in bright, but cloudy weather is recommended. If possible, the cameras' white balance levels as well as their apertures and shutter speeds should be consistent and therefore configured manually – especially when a mostly bright route contains some dark areas like tunnels or forests.

For the sound simulation, we recommend using the audio track recorded by the center camera or by an additional microphone mounted inside the car during the video recordings as a basis. Additional sounds – e.g., the sound of closing doors, noises of other passenger or signal sounds – can be digitally created or recorded separately and merged afterwards. For example, for the second experiment some extra sounds from busses and trams were recorded to simulate the context of a public transport environment. Voice prompts – e.g., announcing upcoming stops – can be recorded with sufficient quality by using built-in microphones of smartphones or laptops.

3.1.3 Postprocessing Video and Audio

To create the immersive video, we postprocessed the three videos with Adobe After Effects (AE) CC 2019. The videos were placed in a virtual three-dimensional room within AE resembling the dimensions of the office room (Figure 3.1, 3.2) to adjust their overlapping areas, perspective, scaling, distortion (to remove the fish eye effect) and position. This enabled us to precisely synchronize and calibrate the footage and to correct minor flaws of the recording process. As the footage was recorded in 2K (and the projectors were only capable of displaying 1080p), it provided sufficient video quality to make the adjustments. The sounds were normalized using Adobe Audition CC 2019 and distorting noises were removed. Furthermore, some extra environmental sounds, signal sounds and voice prompts for our second study were added.

3.1.4 Synchronized Playback

All components, i.e., three video files, one sound file and (optional) HMI displays needed to start synchronously. The best operational setup we tested consisted of a single Macbook Pro (2015) controlling both, sight and sound simulation. Therefore, all video and audio files were opened in QuickTime Player (v. 10.5) and an Apple Script was used to trigger the play event for all opened files.

3.2 Expert Study

To explore the proposed setup, its strengths and weaknesses, HCI professionals were invited. Similar to conducting expert-based usability evaluations of in-vehicle systems [108], we wanted the experts to share their unbiased opinions on the simulation and to provide insights on how to improve the setup. The main purpose of study 1 was to investigate the quality of the simulation, to eliminate potential issues with the setup and to derive optimizations.

3.2.1 Participants

Nine participants (4 female, 5 male, 0 diverse, 0 n/a) with an average age of $M = 29.56$ years ($SD = 4.22; min = 23, max = 38$) took part in the study. They were recruited internally from design and development teams (but were external to the project) and had a professional background in HCI.

3.2.2 Experimental Design and Procedure

After a short introduction from the experimenter, the participants filled out a demographic questionnaire. Then, they took a 20 minutes ride in the AV simulator while thinking-aloud, i.e. they verbalized any thoughts they had during the ride. Directly after the ride, participants filled out a questionnaire to evaluate the quality of the simulation and the experienced level of realism. We chose the Igroup Presence Questionnaire (IPQ) [224, 225] for a standardized assessment of presence as it provides "the highest reliability within a reasonable timeframe" [229] among presence questionnaires. In the last phase, the experimenter conducted a semi-structured interview with the participant to learn about their experiences in the simulator setup and to uncover issues and optimization potential.

3.2.3 Results

Results of the IPQ show positive ratings in the four subscales (Table 3.1). In terms of the scale Experienced Realism the results are, however, slightly on the lower side of the scale. The results of the IPQ are backed by ratings for the single item "I

Table 3.1 – $M(SD)$ of the IPQ [224, 225] subscales [from 0 = low to 6 = high] from study 1 ($N = 9$).

Subscale	M	SD
Spatial Presence	4.00	0.51
Involvement	3.36	1.05
Experienced Realism	2.93	0.58
General	3.56	1.33

found the ride in the simulator realistic." with $M = 3.56$ ($SD = 0.53$) (scale range from 0 = not at all to 5 = fully). Qualitative feedback from the participants of the exploratory study supports the findings regarding presence perception and the general suitability and potential of the AV simulator. All nine experts commented positively on the context-based prototyping approach. Potential for optimization was in particular revealed regarding the sound simulation. Participants suggested to add sounds of opening and closing doors, noises from other passengers and signal sounds for announcing the next stop. Furthermore, three participants emphasized the idea of increasing presence perception by adding actors to simulate passengers getting on and off during shared rides.

3.3 User Study

In study 2, we intended to evaluate the setup with a larger sample and to find out whether AV simulator studies would actually benefit from involving an actor mimicking the behavior of other passengers in terms of participants' subjective presence perception, technology acceptance and motion sickness. Furthermore, the simulator was extended by including the proposed additional sounds. As other passengers are omnipresent in public transportation (and shared rides) and therefore an important part of the context, study 2 aimed to investigate their effects on both, the simulation and overall technology acceptance of shared AVs. However, the presence of others might induce stress resulting in adverse effects on passengers' wellbeing [79]. We expected the negative effects of involving an actor to be rather small and hoped to increase participants' presence perception within the simulation. Furthermore, we expected to discover positive changes in participants' acceptance regarding the use of a shared AV. Consequently, the following hypotheses for study 2 were derived:

H1 The involvement of an actor has a positive effect on participants' presence perception in AV simulator rides.

H2 The involvement of an actor has a negative effect on participants' wellbeing in AV simulator rides.

H3 The involvement of an actor has a positive effect on participants' technology acceptance of shared AVs.

3.3.1 Participants

To achieve sufficient power ($> .80$) with an alpha error of $\alpha \leq .05$, a required sample size of $N_{\text{a-priori}} = 27$ was calculated using G*Power for Mac (v. 3.1.9.4). Medium effects according to Cohen [46] were assumed due to practical considerations, e.g., economic viability, as smaller effects might not warrant the increase in setup effort by enlisting an actor. Thirty-one participants (15 female, 16 male, 0 n/a) with an average age of $M = 31.97$ years ($SD = 10.46; min = 18; max = 54$) took part in study 2. Thus, an actual power of .859 was achieved. All participants were recruited via online postings and received financial compensation. 58 % of participants were holding a university degree and an additional 26 % had a higher secondary school leaving certificate. The affinity for technology interaction (ATI) score [94] of $M_{\text{ATI}} = 4.41$ ($SD_{\text{ATI}} = 0.76; 1 = low; 6 = high$) indicates high technology affinity among the sample.

3.3.2 Experimental Design

The study used a counterbalanced within-subjects design with a within-subjects factor of riding with an actor or not. To avoid systematic carry-over effects, condition order was pseudo randomized, ensuring an equal number of orders. Dependent variables (presence perception, well-being, technology acceptance) and their respective operationalization are listed in Table 3.2.

Again, we used the IPQ in combination with the single item 'feeling of reality' („I found the ride in the simulator realistic.") to assess presence perception. To evaluate participants' wellbeing and corresponding adverse effects (e.g., simulator sickness) a German translation of the Motion Sickness Assessment Questionnaire (MSAQ) [101] along with the single item 'feeling of comfort' ("I felt comfortable during the ride.") was used. Regarding H3, a German translation [138] of the Acceptance Questionnaire by

Table 3.2 – Dependent variables and their corresponding operationalization for study 2.

Factor	Operationalization
Presence perception	Igroup Presence Questionnaire (IPQ) [224, 225] ’Feeling of reality’ – single item "I found the ride in the simulator realistic."
Wellbeing	Motion Sickness Assessment Questionnaire (MSAQ; German translation) [101] ’Feeling of comfort’ – single item "I felt comfortable during the ride."
Technology acceptance	Acceptance questionnaire of Van der Laan et al. [256] in German translation [138]

Van der Laan et al. [256] was applied. In addition to the quantitative measurements, the AV simulator and the prototype of a passenger information system were examined exploratively by observing the reactions and behavior of participants during simulator rides and interviewing them afterward to gather qualitative feedback.

3.3.3 Procedure

On arrival, participants received a short briefing on the study including general information on (shared) AVs, the general objective of the study, and information on simulator sickness. Furthermore, they signed a participation consent form.

Participants took two rides in the AV simulator. Before each ride, the scenario (taking a shared AV to a park and back) was presented and a paper ticket was given to the participants. Each ride took about 14 minutes. In both rides, the simulated SAV stopped twice before the participants reached their destination (Figure 3.5). Another passenger joined the ride at the first stop and left at the second stop. An information display, as well as the sound simulation (step noises), provided information about another passenger getting on/off the vehicle. In one of the two rides (randomly permutated) an actor representing this passenger physically entered the AV simulator. Participants did not receive a briefing on this prior to the rides. In the condition without an actor, the other passenger entered only ‘virtually’ (i.e., he was only represented by the sounds being heard). In both conditions, the other passenger’s getting on/off was displayed on an information display (Figure 3.6). At the third ‘end’ stop participants reached their target destination. After each ride, participants filled out a digital questionnaire to assess the variables listed in Table 3.2. At the end of the session, they received a debriefing and their compensation.

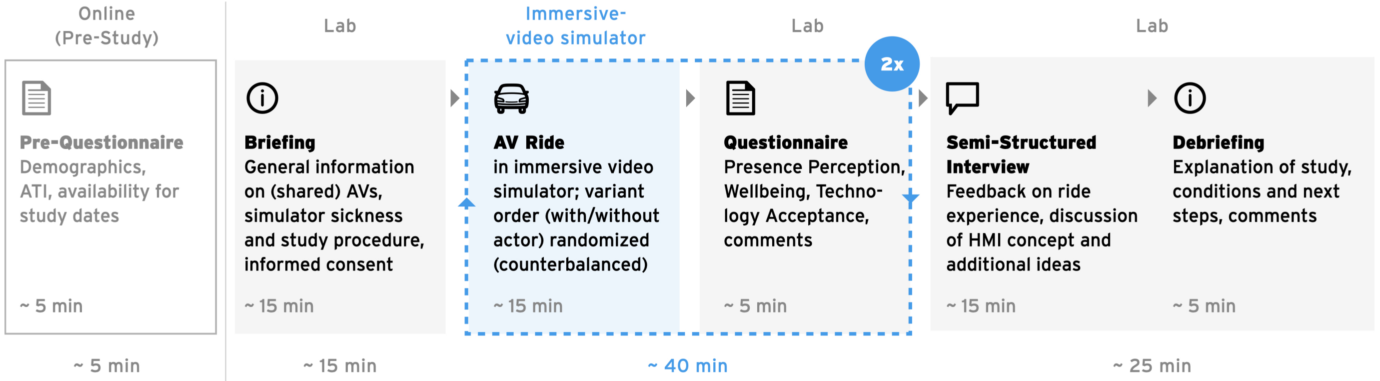


Figure 3.4 – Flow chart of study 2’s procedure.

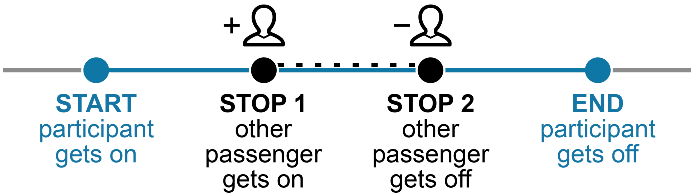


Figure 3.5 – Schematic illustration of the ride sequence.

3.3.4 HMI Concept

During the ride, an audiovisual HMI (26-inch display and stereo sound system) communicated the SAV's current location (position in the map), upcoming stops, the planned route and traffic conditions (e.g., delays caused by congestion). For study 2, some parts of the visual information display (Figure 3.6) were personalized, e.g., respective passenger destinations were indicated via unique ticket IDs. The visual information was complemented with signal sounds and voice prompts announcing upcoming stops. The audiovisual HMI was integrated in the AV simulator as a video-based prototype (Figure 3.1).

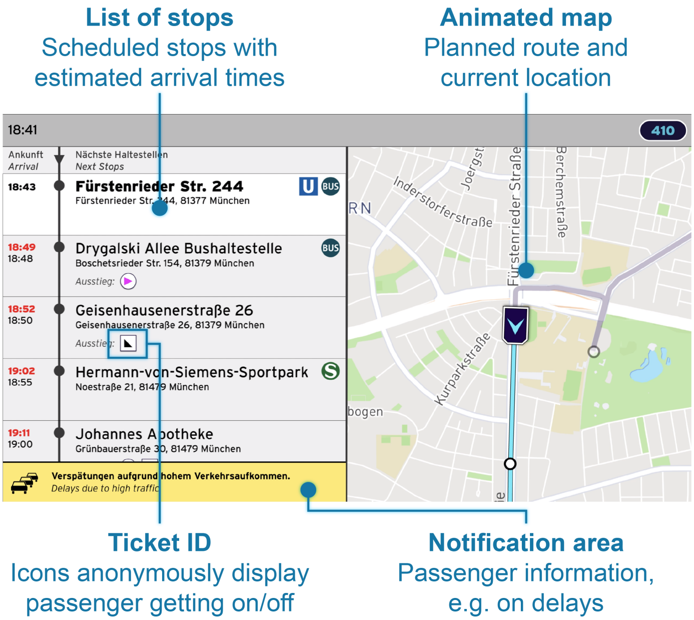
3.3.5 Results

Tests on normality (Shapiro-Wilk) were performed on the underlying distributions prior to the statistical analysis. In case they returned non-significant, parametric inferential statistics (paired-samples t-tests) were calculated. Otherwise, Wilcoxon signed-rank tests were computed. For the statistical analysis JASP for Mac (v. 0.10.2) was used. The reported participant statements were translated into English by the authors.

Presence perception

Descriptive statistics and plots of the IPQ scales (Figure 3.7) reveal similar results to study 1. A tendentially positive evaluation of presence perception within the AV simulation, especially in terms of spatial presence and experienced realism, is observable. Regarding the Involvement subscale, a slight trend in favor of the condition without actor is recognizable. Inferential statistics (paired samples t-tests) on the IPQ subscales corroborate these observations but do not return significant results (Table 3.3).

The ratings of the single item 'feeling of reality' are generally high in both conditions



18:41

Ankunft *Arrival*	Nächste Haltestellen *Next Stops*
18:43	**Fürstenrieder Str. 244** Fürstenrieder Str. 244, 81377 München
18:49 18:48	Drygalski Allee Bushaltestelle Boschetsrieder Str. 154, 81379 München *Ausstieg:* ▶
18:52 18:50	Geisenhausenerstraße 26 Geisenhausenerstraße 26, 81379 München *Ausstieg:* ◣
19:02 18:55	Hermann-von-Siemens-Sportpark Noestraße 21, 81479 München
19:11 19:00	Johannes Apotheke Grünbauerstraße 30, 81479 München

Verspätungen aufgrund hohem Verkehrsaufkommen.
Delays due to high traffic

Figure 3.6 – HMI concept (©2020 by Ergosign GmbH; map: ©Mapbox, ©OpenStreetMap) providing information on the current location of the SAV (position in map), the planned route, passengers getting on/off, and traffic conditions.

(Table 3.4) and higher than in study 1. This is also strongly supported by anecdotal evidence (i.e., by the reports of the participants in the questionnaire and in informal talks during and after the experiment).

Eleven participants commented positively on the feeling of reality. For example: *"incredibly real."* (P8); *"I had the feeling of actually sitting in a car"* (P11); *"[it feels] very real, although one sits on a chair in a room."* (P15); *"the ride reminds me of a normal car or bus journey. It was very real"* (P17); *"the ride was very realistic due to the environment and the people."* (P17); *"the immersion is extremely good due to the real pictures" (P19); "comparable to reality"* (P23). Three participants (P9, P11, P12) actually reported that their body wanted to move in accordance with the visual simulation (e.g., when the car stopped or accelerated). Five participants (P11, P17, P23, P24, P26) explicitly appreciated the conservative and anticipatory driving style of the simulated AV.

Six participants, however, commented negatively regarding the feeling of reality: *"movements of the chairs and the simulator itself are missing"* (P13); *"design of the simulator's interior feels more like a waiting room"* (P13); *"Although the situations were good represented,*

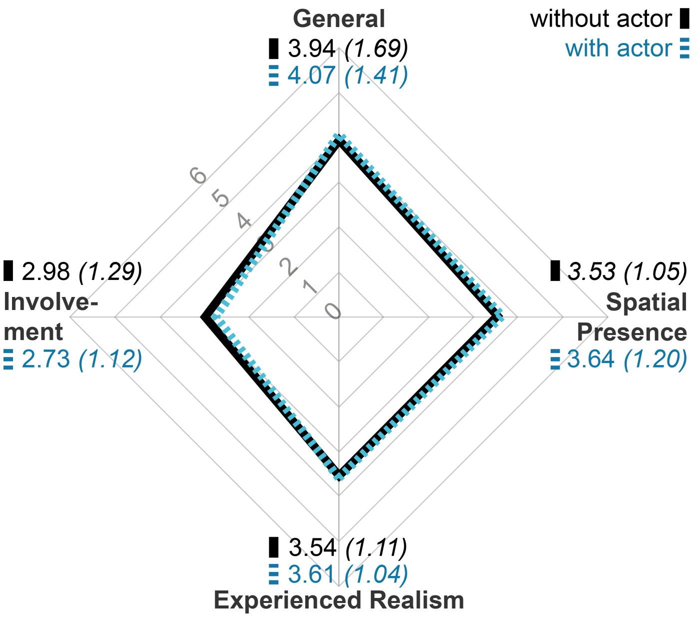


Figure 3.7 – $M(SD)$ of the IPQ [224, 225] subscales [from 0 = low to 6 = high] for the two conditions of study 2 (without actor – with actor; $n = 31$).

I noticed the equipment" (P15); *"I always knew inside that it is fake"* (P14); *"the left image was much sharper than the other videos"* (P21); *"it doesn't feel so realistic when there is always a free parking slot available [for the AV] at just the right place"* (P26); *"color variances in the video projections had a negative effect on the ride experience"* (P31).

In contrast to the results of the IPQ's subscales, results of the single item 'feeling of reality' (Table 3.4) show a positive tendency toward using an actor. A paired samples t-test exposes a significant difference between the conditions ($t_{\text{paired}}(30) = -2.64, p = .007$) favoring using an actor with a medium effect of $d_{\text{paired}} = -0.47$. Seven participants explicitly support the measurements with their comments, e.g., *"It was kind of spooky when the guy came in."* (P8); *"it felt more real when someone entered the vehicle"* (P10); *"It felt really realistic. I was almost shocked when the person entered the vehicle."* (P11); *"the entrance [of another passenger] was very realistic"* (P17); *"the co-passenger made the ride more realistic"* (P18).

Table 3.3 – Paired samples t-tests and Cohen's d on the IPQ [224, 225] subscales. Hypothesis is without actor rating decreases against with actor.

Subscale	t_{paired}	df	p	d_{paired}
Spatial Presence	-0.826	30	.208	-0.148
Involvement	1.598	30	.940	0.287
Experienced Realism	-0.356	30	.362	-0.064
General	-0.626	30	.268	-0.112

Table 3.4 – $M(SD)$ for the rated single item "feeling of reality" [from 1 = not at all to 5 = fully] for the two conditions (without actor – with actor; N = 31).

Subscale	Without actor	With actor
Feeling of reality	$3.81(1.01)$	$4.23(0.85)$

Technology Acceptance

Both subscales of the acceptance questionnaire [256] show a slight positive trend toward involving an actor (Table 3.5). Paired-samples t-tests do not reveal significant differences (Table 3.6). However, the tests indicate a non-significant trend with a small effect of $d_{\text{paired}} = -0.28$ regarding the usefulness ($t_{\text{paired}}(30) = -1.56, p = .065$) in favor of actor involvement.

Table 3.5 – $M(SD)$ of the acceptance questionnaires' [256] subscales [from -2 = negative to 2 = positive] for the two conditions (without actor – with actor; $n = 31$).

Subscale	Without actor	With actor
Usefulness	$1.17(0.51)$	$1.25(0.49)$
Satisfying	$1.23(0.52)$	$1.27(0.53)$

Wellbeing

Whilst descriptive statistics of the MSAQ generally show low values in all subscales, slightly higher values are observable regarding the condition involving an actor (Table 7). Wilcoxon signed-rank tests reveal significant differences in the subscale Gastrointestinal ($W = 50.00, p = .038, N = 31$) with a medium effect of $r_{\text{rb}} = -0.47$, as well as in the Overall MSAQ scale ($W = 65.00, p = .024, N = 31$) also with a medium effect of $r_{\text{rb}} = -0.49$. The results indicate significantly higher values for motion sickness in SAV simulator rides involving an actor.

Table 3.6 – Paired samples t-tests and Cohen's d for the Acceptance questionnaire [256] subscales. Hypothesis is without actor rating decreases against with actor.

Subscale	t_{paired}	*df*	p	d_{paired}
Usefulness	-1.56	30	.065	-0.28
Satisfying	-0.63	30	.265	-0.11

Table 3.7 – $M(SD)$ of the MSAQ [101] subscales [from 11 = low to 100 = high] for the two conditions (without actor – with actor; $n = 31$).

Subscale	Without actor	With actor
Gastrointestinal	17.12(9.73)	22.04(16.67)
Central	17.06(10.76)	19.28(15.13)
Peripheral	13.50(6.45)	13.86(6.62)
Sopite-related	20.61(11.89)	22.04(13.07)
Overall	17.29(6.94)	19.65(9.81)

The subjective feeling of comfort (*"I felt comfortable during the ride."*) achieves high values in both conditions (Table 3.9) with no relevant difference ($W = 44.00, p = .672, N = 31$). Four participants explicitly mentioned symptoms related to simulator sickness: *"I got a little woozy, but I'm fine"* (P2); *"when I turned around, my head was slightly spinning"* (P8); *"I felt uncomfortable driving over the cobblestone at the end of the second ride"* (P13); *"sometimes I got a little nauseous during the ride"* (P15).

3.4 Discussion

We evaluated the presented AV simulator in two empirical studies. Both studies investigated the suitability of the simulator for context-based prototyping and evaluation. The findings of the expert consultation in study 1 were primarily used to discover issues and optimization potential of the setup. In study 2, we focused on investigating the impact of involving an actor in AV simulator studies in terms of participants' presence perception, wellbeing and technology acceptance.

Table 3.8 – Wilcoxon signed-rank test and rank-biserial correlations for the MSAQ [101] subscales. Hypothesis is without actor rating decreases against with actor. *significant ($p < .05$)

Subscale	W	p	r_{rb}
Gastrointestinal	50.50	.038*	−0.47
Central	30.00	.083	−0.43
Peripheral	2.50	.231	−0.50
Sopite-related	62.50	.163	−0.27
Overall	65.00	.024*	−0.49

Table 3.9 – $M(SD)$ for the rated 'feeling of comfort' [from 0 = not at all to 5 = fully] for the two conditions (without actor – with actor; N = 31).

Subscale	Without actor	With actor
Feeling of comfort	4.07(0.89)	4.03(0.98)

3.4.1 Immersive Video-Based AV Simulation

The tested prototype received quite positive ratings in both studies regarding presence perception. Considering subjective quantitative and qualitative ratings in terms of participants' presence perception and feeling of reality, the results are encouraging. The improvements of the sound simulation based on the findings of study 1 (e.g., including sounds of open/closing doors, other passengers and signal sounds) seem to have a positive effect on the quality of the simulation and should therefore be further investigated. Extending the video-based setup with a motion simulation might increase presence perception. As a participant mentioned in study 2, rides might possibly feel more real when they are less smooth. For example, the AV might stop only close by to a certain scheduled stop, but not exactly at the given location. In general, the results provide support for the suitability of the method to enable straightforward context-based design and evaluation.

3.4.2 Involving an Actor in Shared AV Simulation

Anecdotal evidence and a significant medium effect in the single item 'feeling of reality' suggest a positive influence of actor involvement. However, in contrast to our expectations, this is not backed by the results of the standardized IPQ. Moreover, while the general feeling of comfort was rated high in both conditions with no relevant difference, negative effects of actor involvement on participants' well-being have been revealed by the MSAQ. Although, the overall occurrence of motion sickness symptoms measured by the MSAQ was very low, they were slightly, but significantly higher when an actor was involved in the simulation. This might have been caused by a disruption of participants' immersion when the actor entered/left the simulation. It might also reflect a general feeling of stress and discomfort when unknown people are present (see [79]). Despite a non-significant trend with a small effect in the usefulness subscale (indicating slight differences favoring the involvement of an actor), we did not observe a statistically relevant effect regarding participants' acceptance ratings of (shared) AVs. On this basis, no conclusion can be drawn regarding neither a positive nor a negative effect of actor involvement on technology acceptance. To sum up, Hypothesis H1 (positive effect of actor involvement on participants' presence perception) is partly supported and Hypothesis H2 (negative effect of actor involvement on participants' wellbeing) is supported by the results, whereas, Hypothesis H3 (positive effect of actor involvement on participants' technology acceptance) is not supported.

3.4.3 Challenges and Limitations

The proposed AV simulator provides a simple framework for creating high-fidelity prototypes of (S)AVs. Some challenges should, however, be considered when using the method. As cameras are mounted behind the windows to capture the footage, it is only possible to create visual simulations under appropriate weather conditions, as for example raindrops or reflections might restrict the visibility. In order to create suitable simulations, careful advance planning of scenarios is required because editing of existing footage is only possible within tight limits.

Since the simulation is based on videos created during driving in public, undesired artifacts (e.g., caused by the behavior of other road users, camera focus or orientation) may occur and have adverse effects on quality and precision of the simulation. Furthermore, controllability of the video-based simulation is limited, especially in comparison to computer-generated environments. Regarding the reported studies, limitations are primarily induced by the sample composition and the questionnaires used. The ATI indicates high technology affinity among the well-educated participants of study 2, which is considered a common phenomenon in HCI research [94] and might impair external validity. The used IPQ was initially created to measure the subjective sense of presence in virtual environments [225].

Since the created AV simulator setup differs from ‘classic’ computer-generated virtual environments, it cannot be directly compared to provided benchmarks, restricting the interpretability of results.

3.4.4 Future Work

The immersive video-based AV simulator provides a suitable basis for context-based prototyping and evaluation of interfaces for (shared) AVs. Further studies might, for example, use the setup for usability testing of in-vehicle HMIs or mobile apps, but also for user research (e.g., regarding technology acceptance and trust). In addition, the simulator might be used by designers and researchers to support research-through-design approaches, e.g., for ideation techniques like body-storming (or ‘car-storming’ as described by Krome et al. [145]). Furthermore, the video-based approach could be transferred to other domains and used to investigate experiences in other future modes of transport (e.g., autonomous air taxis).

Regarding the investigation of the potential effects of actor involvement, further studies can build upon the findings of study 2 and extend the operationalization of the independent variable, e.g., by adding a third condition where there is no other passenger present at all – neither physically (as an actor) nor virtually (in terms of sound or imagery).

Since the described approach is limited regarding the controllability of the video-based simulation (i.e., the immersive video), further work should put special attention toward refining and potentially standardizing the way of manual data collection, while maintaining the approach’s simplicity. Quality and precision of the video recordings might, for example, benefit from video rigs (like e.g., used by Gerber et al. [98]), fixed camera mounts or by using a single 360° camera instead of multiple cameras. However, using special equipment would also make the setup less simple and more expensive. The immersive video might also be extended by CGI (e.g., [98]) or combined with real-world data-based generation of virtual environments (e.g., [13]) to simulate specific situations. Researchers and designers might also share and exchange audio and video files in order to minimize qualitative differences and to conduct comparable studies.

To further investigate the AV simulator’s validity and cost-efficiency, experiments should be compared to both, real-world experiments and laboratory experiments.

To conduct studies in the real world, vehicles with ‘real’ autonomous driving functionalities might be used. But, as mentioned above, such studies are often only feasible within tight limits. Thus, utilizing ‘common’ vehicles in combination with wizard-of-oz techniques might be – depending on scenario and study objective – a better fit. Regarding the comparison to laboratory studies, ‘standard’ setups without contextual simulation should be taken into consideration as well as setups with CGI-based simulation. Altogether, this would allow for a profound evaluation of the simulator's external validity. It would also support a better understanding of the type of prototyping fidelity needed for cost-effective design and evaluation of AV interfaces and, consequently, enable the creation of a methodological framework.

3.5 Conclusion

In this chapter, we presented a simple immersive video-based AV simulator as a prototyping and evaluation method for (S)AVs and AMoD systems. The cost-effective setup consisting of real-world videos and a CAVE-like environment is comparatively easy to (re-)create and can be likewise used by designers, engineers and researchers as a prototyping framework for design and evaluation activities. It can be used to study user behavior and to counteract human factor challenges related to (S)AVs from early development phases on. Presented results of two user studies can be taken as initial evidence for the simulators' suitability for context-based prototyping of HMIs for AVs. However, due to the simple approach, the method is limited in terms of precision and controllability of the simulation. Although, we found some support for the idea of using an actor as a part of the simulation of a shared ride, we did not find a significant positive impact on participants' presence perception. Moreover, it might also have adverse effects on their well-being (e.g., regarding simulator sickness). Thus, we do not generally recommend using an actor in AV simulator studies. Nevertheless, actor involvement might still offer valuable insights for prototyping specific scenarios relying on social context.

3.6 Chapter Summary

In this chapter, we presented a simple approach for context-based prototyping of HMIs for (shared) AVs. Inspired by previous works of Kray et al. [143] and Gerber et al. [98], we applied immersive-video [143] to create a straightforward and cost-effective (shared) AV simulator. We captured the required real-world video and audio footage with three low-budget cameras while driving through urban traffic. The footage was displayed with three video projectors in a common office space. Since AVs are driverless, the setup did not require 'traditional' control elements such as pedals or a steering wheel. Supported by the enclosed area, the projections and walls already created a spatial mock-up.

Although being a simple and rather abstract mock-up, the video-based approach provided a suitable setup for context-based evaluation of HMI concepts. We judge this based on the results of an expert study ($N = 9$) and a user study ($N = 31$) in which we investigated presence perception and simulator sickness. While the setup offers good reliability and can be easily recreated at other places and with other components, the approach has limits in terms of external validity and the degree of control over the simulation.

We supplemented the audio-visual simulator with a social context simulation in the user study with the intention to increase participants' immersion. Here, the simulated shared AV stopped occasionally when another passenger joined and left the ride. We compared two conditions: 1) a ride with only virtual simulation (i.e., sounds and display in the HMI) of a passenger getting on and off the shared ride and 2) a ride with an actor physically entering and leaving the simulator. While participants assessed their 'feeling of reality' as significantly higher with the actor, the standardized IPQ presence questionnaire did not support the finding. Moreover, adverse effects on participants' well-being in terms of higher occurrence of motion sickness symptoms were revealed. While this might have also been due to a general feeling of discomfort when other (unknown) people are present, we concluded that actor involvement does not affect participants' presence perception in a way that would justify the extra effort. However, based on our learnings in Chapter 6, we want to point out that actor involvement can indeed make sense depending on the scenarios prototyped and the importance of the social context.

In terms of **RQ2**, this chapter contributes a simple and cost-effective approach for context-based prototyping using immersive-video. We presented and demonstrated how it can be used to create a lightweight simulation for contextualized prototyping of in-vehicle HMIs for (shared) AVs. The conducted studies provide initial evidence for the approach's suitability and show how it can be combined with social context simulation through actors. A massive benefit of the approach is that no specific skill set, such as programming skills, is required to create the simulation. As a result, the approach can easily be applied, adapted (e.g., by using TVs instead of video projectors like we show in Chapter 5), and extended by both researchers and practitioners aiming to consider contextual factors in their design activities while also accounting for potentially limited resources (e.g., in terms of budget, time, or development skills).

4

Prototyping Augmented Reality Windshields with Wizard-of-Oz

Wizard-of-Oz (WoOz) is a prototyping method used to test systems that are not (yet) available or feasible. Therefore, study participants are made believe that they are interacting with an intelligent and/or automated system while humans do in fact simulate it – the so-called *wizards* [23]. When using the method to prototype AVs, study participants need to believe that the vehicle is driving automated while a hidden human driver – the driving wizard – controls it [19].

In this chapter, we set up a straightforward WoOz setup to investigate whether passengers' acceptance and UX of AVs can be increased by providing transparent information on the AI system's reasoning with the (computer-vision-based) visualization of detected objects and how this information should be displayed during the

---

THE CONTENTS OF THIS CHAPTER WERE PUBLISHED PREVIOUSLY AS PART OF:

[88] **Lukas A. Flohr**, Joseph S. Valiyaveettil, Antonio Krüger, and Dieter P. Wallach. 2023. Prototyping Autonomous Vehicle Windshields with AR and Real-Time Object Detection Visualization: An On-Road Wizard-of-Oz Study. *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. `https://doi.org/10.1145/3563657.3596051`

ride. To answer the corresponding questions, we conducted an on-road WoOz study in a real, urban environment and compared three HMI variants: 1) a baseline concept without object detection against two forms of real-time visualizations of detected objects: 2) a status bar with counts of detected objects per class, and 3) augmented reality overlays.

By combining the WoOz setup, real-time object detection algorithms and augmented reality, we demonstrate a prototyping approach for computational interaction with AVs in the spirit of human-centered AI.

4.1 On-Road Wizard-of-Oz Study

Related work reveals the potential of providing explanations and transparent system feedback to increase acceptance, trust, and UX of automated vehicles [47, 48, 82, 156, 142, 160, 185, 270]. Most of these previous studies (except [185]) investigate vehicles with lower levels of driving automation where a human driver is still required (i.e., up to SAE level 3) and focus on providing system feedback for specific situations (e.g., maneuvers in ambiguous situations [271]). Furthermore, they were conducted online [47, 48], in labs with simulated artificial environments [82, 142, 156, 160, 270], or on restricted (in-door) test tracks [185], but not on real urban roads. The questions arise about whether previous findings from lower automation levels can be transferred to driverless AVs and to dynamic and complex urban real-life environments, and whether, when, and how transparent information and explanations should be displayed in AVs. We address the identified research gap by investigating the following questions in an empirical user study. Q4.1 and Q4.2 contribute to answering our general research question RQ1, while Q4.3 adds to RQ2.

Q4.1 Can we increase AV passengers' acceptance and UX by providing transparent system information via (AR-based) visualization of detected objects in the vehicle windshield?

Q4.2 How and when should this information be displayed during AV rides in urban environments?

Q4.3 How can we create a suitable prototyping framework to investigate Q4.1 and Q4.2, as well as related questions in complex urban real-life environments?

Figure 4.1 – On-road wizard-of-oz prototyping in an electric minivan (i) with a TV mounted on the headrests of the front seats (iii) that displays a video stream from the vehicle's windshield with AR-based real-time object detection visualization (ii) provided by an embedded computing platform (Nvidia Jetson Nano).

4.1.1 Study Design and Prototyping Framework

We adopted a within-subjects design to achieve high internal validity and to minimize the effects of random noise [36], e.g., caused by varying environmental factors. To investigate the effects of real-time object detection visualization on passengers in a natural urban environment (i.e., in real traffic), we created a contextualized prototyping framework based on a WoOz setup in combination with a prototype for a futuristic windshield HMI implemented on an embedded computing board (Nvidia Jetson Nano). Before conducting the study, its design, setup, procedure, and data collection were assessed by the Ethical Review Board of the faculty of Mathematics and Computer Science at Saarland University with the process number 21-11-4. The board did not raise any ethical concerns. Furthermore, the study was conducted in accordance with applicable ethical principles stated in the Declaration of Helsinki [275].

Wizard-of-Oz Setup and Prototyping Considerations

Due to the reasons elaborated in Section 2.3, we opted for a WoOz approach as a basis for the study and the investigation of the derived questions. Inspired by previous works, especially by Karjanto et al. [131] and Detjen et al. [64], we created a straightforward WoOz setup (Figure 4.1 and 4.2) that we used as an on-road simulation of an AV ride through the city. An electric minivan (Mercedes-Benz EQV) served as a basis for the setup. The car came with a modern appearance and offered sufficient space

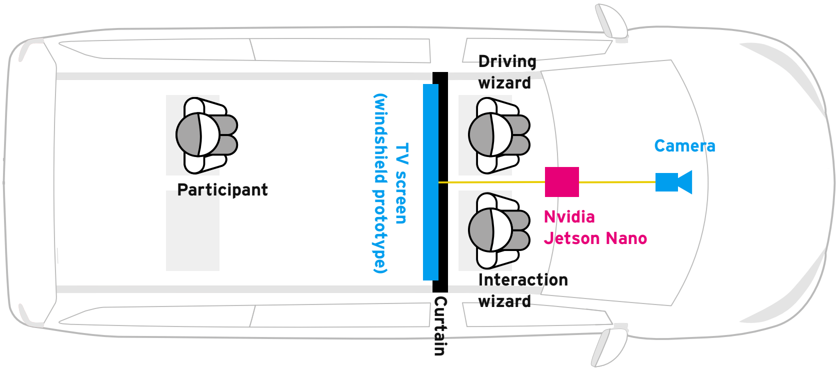


Figure 4.2 – Schematic illustration of the used wizard-of-oz-based prototyping setup.

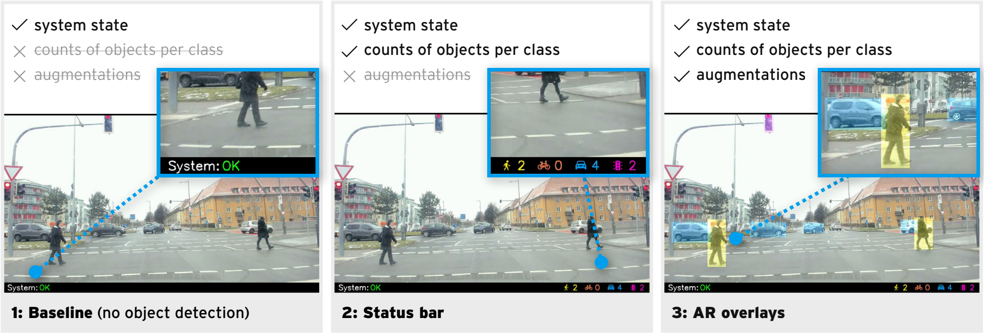
for the setup. Since we used a rental car, we aimed – in contrast to previous works – to create an easily deconstructable setup without the need for physical adjustments (e.g., drill holes) that can also be easily reproduced in similar vehicles. To achieve this, we mounted a TV (Hisense 43" 4K) at the front seats' headrests using a wooden board with a standard TV wall mount and screw pipe clamps. To provide the basis for the investigation of Q4.1 and Q4.2 and as a potential answer to Q4.3, we connected the TV to an embedded computing platform (Nvidia Jetson Nano). The Jetson displayed the HMI prototype (Section 4.1.1), including the video stream of a consumer webcam (Logitech BRIO) mounted in the vehicle's windshield. We then mounted a black curtain with heavy-load magnets and duct tape on the car's ceiling to separate the vehicle's front and back parts. A power inverter (NDDI 600 W) inverted the vehicle's 12 V DC power plug to 240 V AC to power the TV and the Jetson. For safety measures, we added an additional socket with surge protection. Based on the consultation of an automotive expert witness auditing company, we made some final adjustments and optimizations by better securing the load and setup. Lastly, we added a foiling to the vehicle's exterior that marked the vehicle as a research vehicle to support the WoOz cover story used.

Following the recommendations of related work, we instructed the wizard to perform a conservative and relaxed driving style, like "a professional limo driver" [11]. To increase objectivity, all sessions were driven by the same experienced driver who was familiar with the vehicle (familiarization time of 3 weeks prior to the study) and aimed to reproduce the same driving style throughout all sessions consistently. As the WoOz setup limited the view out of the vehicle, the co-driver (interaction wizard) supported the driver in difficult situations during the test rides, e.g., by spotting vulnerable road users when turning right.

Windshield HMI Prototype

The futuristic (AR-capable) windshield HMI prototype was implemented as a graphical user interface (GUI) application displaying image frames from a webcam using OpenCV. Depending on the concept variant – the prototype draws real-time AR bounding box visualizations over detected objects and/or shows a descriptive status bar with counts of objects per class for the detections (Figure 4.3). The detector uses a pre-trained model (YOLOv4 [26] trained on the COCO dataset [155]) optimized for inference using ONNX and TensorRT and runs on the Jetson's GPU. Reducing the complexity of the HMI and the study, the application merges object classes from the dataset into four main headers: pedestrians, cyclists, vehicles, and traffic signs. To reduce latency and jitter from object visualizations and increase the frame rate of the video feed, we implemented a periodical switch to a lower overhead object tracker that was periodically re-initialized by the object detector. The application was implemented on an Nvidia Jetson Nano embedded-computing board with a 4-core CPU, 4 GB RAM, and a 128-core GPU and displayed on the TV. We applied several optimization measures to display the video feed and the object visualizations with a fluent frame rate and sufficient resolution (TensorRT optimizations, joint detection and tracking). This resulted in a feasible resolution of 1280 x 720 pixels at about 24 fps, that was, with regards to the passengers' viewing distance of about 160 cm, sufficient. We want to note that the early computer-vision-based prototype's performance has limits and is not up to the accuracy and precision of cutting-edge sensing systems, e.g., [206, 264]. Nevertheless, the implementation provides a suitable and flexible prototyping basis to investigate our research questions at an early development stage.

For the design of the AR-based object visualizations, we adopted two-dimensional bounding box overlays as they are widely used in the computer vision domain for basic object annotation (Figure 4.3: 3). Depending on the object class (e.g., pedestrian) , the overlays had different colors (e.g., yellow). In the design phase, we also considered approaches and visualization techniques, such as 3-dimensional AR markers or as representations on a separate display, as well as combinations with "classic" information, feedback, and navigation concepts (e.g., displaying the planned route on a map). However, the reported study was intended as an early concept study, which is why we focused on (AR-based) object visualizations. To investigate their



1: Baseline (no object detection)

2: Status bar

3: AR overlays

Figure 4.3 – Overview of the three concept variants displayed in the windshield prototype.

general potential, we created a baseline variant of the prototype without feedback on detected objects (Figure 4.3: 1). Since related work pointed out that the amount of displayed information might affect passengers' experience [185], we created an intermediate variant which visualized detected objects as counts per object class in a status bar only (Figure 4.3: 2). The variants are designed sequentially. I.e., variant 3 also includes the status bar of variant 2. Furthermore, all three variants displayed general information on the overall system state ("System: OK"), which provided passengers with baseline information on the system's functionality throughout the variants. We opted to provide this baseline information for two reasons: to inform (and convince) passengers that the simulated AV is driving autonomously and to ensure them that everything is fine – even if there is no further information displayed. In the conducted study, the system state never changed.

4.1.2 Participants

With a sample of $N = 30$ participants (14 female, 16 male, 0 diverse, 0 n/a) between the ages of 20 and 70 ($M = 37.6, SD = 11.9$), we achieved a statistical power of $.84$ (calculated with G*Power 3.1) for the calculation of inferential statistics (repeated measures analysis of variance (RM-ANOVA) with within factors and three measurements) assuming medium effects according to Cohen [46] and an alpha error rate of $\alpha \leq .05$. The sample had a medium-high affinity for technology interaction (ATI-S [268]: $M = 4.4, SD = 1.3; min = 1, max = 6$) and was well educated (highest degree: 19 with university degrees, six with advanced school-leaving certificates, three with intermediate school-leaving certificates, two with other degrees). Three participants had an uncorrected visual impairment (two myopia, one red-green color blindness), which they reported not having posed a problem during the study. All participants were external from our institution and recruited via online postings, mailing lists, and advertising posters. Each participant received financial compensation of 30 €.

4.1.3 Procedure and Data Collection

Each participant took part in an individual session with an experimenter and a note taker, who also took over the roles of the driving and interaction wizards during the test rides. The sessions took about 90 min and were structured into three main phases (Figure 4.4): 1) briefing and pre-questionnaire in the lab, 2) test rides and consecutive questionnaire in the WoOz vehicle, and 3) semi-structured interview and debriefing in the lab. Following a mixed-method approach [52], we collected both qualitative and quantitative data. For an in-depth post hoc analysis, we recorded audio during the rides and the interviews and took notes during the sessions. For the quantitative assessment and comparison of the HMI variants, we used standardized UX, trust, and acceptance questionnaires and single-items to assess perceived risk, safety, wellbeing, and nausea during the rides (Table 4.1).

Briefing and Pre-Questionnaire

At the beginning of the study session, the participant received a detailed briefing on the study's purpose and procedure. This was already initialized our WoOz deception. As a part of our cover story, we explained the basics of autonomous driving technology and automation levels. We told participants that we would conduct the test rides with an actual AV capable of handling all driving situations but requiring the presence of a safety driver (the driving wizard) due to current legal regulations. Furthermore, we declared that we wanted to evaluate futuristic windshield HMI concepts that are technically not yet feasible to be implemented in the vehicle. This served as the explanation for the TV-based prototyping. By providing passengers with this information, we aimed to shift the focus toward the HMI prototype and away from the WoOz setup. Furthermore, we explained to participants how the AV's object detection works and that AVs use it to navigate safely through traffic. We outlined that some of the tested HMI concepts might provide this sensor information also to passengers to optimize their experience. It was added that the tested concepts are currently in an early prototyping phase and are, thus, using not the actual AV sensors

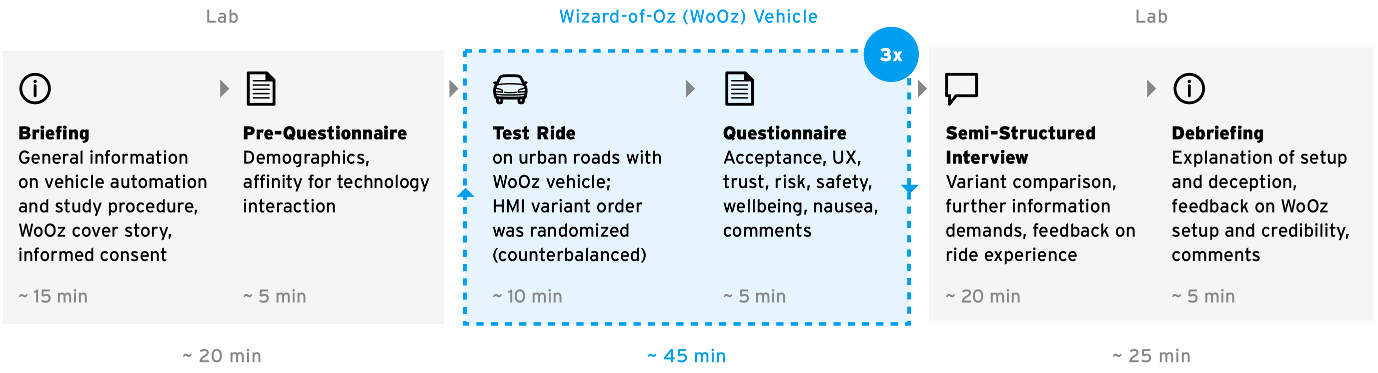


Figure 4.4 – Study procedure of the on-road wizard-of-oz study.

but a single camera that we included in the AV for research purposes. Due to this early development stage, we explained that the system's performance is limited and might affect the correct display of the HMI information. After the briefing, participants signed an informative participation consent form and filled out a pre-questionnaire to provide information on their demographics and affinity for technology interaction (ATI-S [94, 268]).

Table 4.1 – Dependent variables and their operationalization.

	Scales	Items
Acceptance	Satisfying [256]	4 bipolar items; 5-point scale
	Usefulness [256]	5 bipolar items; 5-point scale
	Perceived Enjoyment [41]	3 items; 5-point Likert-type scale
	Intention to Use [41]	2 items; 5-point Likert-type scale
Trust	Trust in Automation [141]	2 items; 5-point Likert-type scale
	Reliability/Competency [141]	6 items; 5-point Likert-type scale
	Understandability/Predictability [141]	4 items; 5-point Likert-type scale
UX	Pragmatic UX [222]	4 bipolar items; 7-point scale
	Hedonic UX [222]	4 bipolar items; 7-point scale
MISC	Risk	1 item; 5-point Likert-type scale
	Safety	1 item; 5-point Likert-type scale
	Wellbeing	1 item; 5-point Likert-type scale
	Nausea	1 item; 5-point Likert-type scale

Test Rides and Questionnaire

The test rides were conducted as a round-trip through an urban environment with two stops at parking lots and about 10 min driving time per variant. The variant order varied (counterbalanced) between sessions to decrease carry-over effects. Before starting the ride, participants were given some final notes on the setup. We encouraged them to think aloud and explained to them once again that they could pause or quit the study at any time without consequences. At the two stops, we changed the HMI variant and asked participants to fill out a digital questionnaire on a tablet to assess the respective HMI variant in terms of our dependent variables (Table 4.1). For the assessment of acceptance, we used the *Satisfying* and *Usefulness* scales of Van der Laan et al. [256] and the scales *Perceived Enjoyment* and *Intention to Use* of Chen's TAM adaption [41]. As related work has identified trust as a key acceptance challenge for AVs [41], participants also assessed the variants in terms of trust and related factors using the scales *Trust in Automation*, *Reliability/Competency*, and *Understandability/Predictability* by Körber [141].

For the assessment of *pragmatic and hedonic UX*, we used the short version of the User Experience Questionnaire (UEQ-S) [222]. In addition, we used single-item scales to let participants assess perceived *risk* ("I considered the ride risky."), *safety* ("I felt safe during the ride."), *wellbeing* ("I felt comfortable during the ride."), and *nausea* ("I felt nauseous during the ride."). Participants could comment on their assessments via free-text input fields. Following the recommendation of [19] to collect environmental data of the test rides, experimenters documented weather conditions and traffic density.

Semi-Structured Interview and Debriefing

After the test rides, we conducted a semi-structured interview using closed and open questions. We recapitulated the rides and HMI variants and talked to participants about what they liked and disliked, their preferences, and what they would suggest for future systems. We also asked participants which variant they liked best and why. At the end of the interview, we lifted the WoOz deception and explained the reasons. After the explanation, we asked participants the *WoOz control question* ("Did you believe that the vehicle was driving autonomously?") to directly assess the deception's effectiveness.

4.2 Results

For the quantitative results, we used *JASP 0.16* [127] and *jamovi 2.2.5* [249] to calculate descriptive and inferential statistics. In a second step, we analyzed the qualitative data from the interviews, ride recordings, and questionnaires. All recordings were transcribed using the speech-to-text function of *Condens* [49], reviewed, and manually optimized afterward. Following an inductive thematic analysis approach [29, 28], three researchers worked collaboratively. We used Condens and a digital *Miro* whiteboard to analyze and structure the data in order to identify patterns that describe essential information concerning our Q4.1 and Q4.2. Each researcher started with analyzing a few sessions and derived an initial set of codes which was then reviewed by the others and merged to create a joint codebook. The codebook and coding fragments were iteratively refined throughout the analysis. Finally, the thematic analysis was complemented with the questionnaire results and session notes.

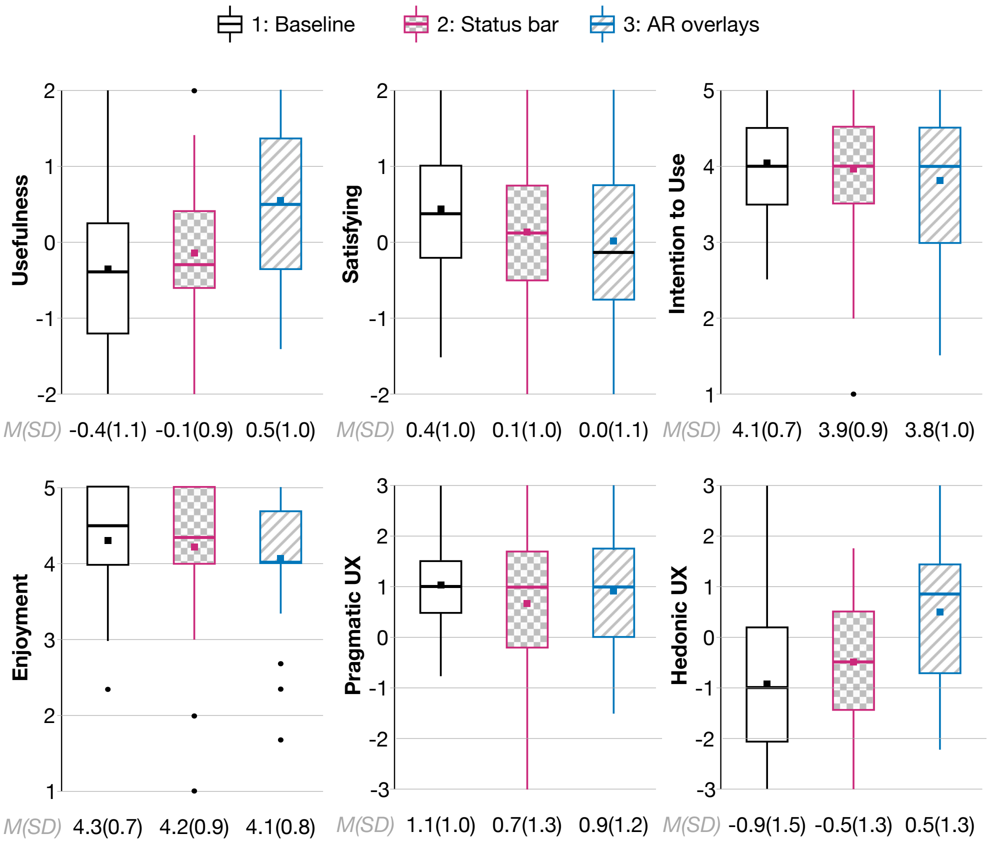


Figure 4.5 – Boxplots and $M(SD)$ of acceptance and UX scales for the three concept variants.

4.2.1 Dependent Variables

Besides a descriptive analysis, we conducted repeated measures analysis of variances (RM-ANOVAs) to search for statistically relevant effects. For the interpretation of calculated effect sizes, we refer to Cohen [46]. If one or multiple assumptions of the RM-ANOVAs (independence, normality, sphericity) was found violated for a particular scale, we calculated non-parametric Friedman tests and Conover's post hoc comparisons.

Acceptance

Results of the VdL acceptance questionnaire [256] show medium ratings of *Satisfying* and *Usefulness* scales with the Baseline achieving highest ratings with regard to *Satisfying* and the AR overlays highest in terms of *Usefulness* (Figure 4.5). While no significant difference was found for *Satisfying* ($F(2, 58) = 1.590, p = .213, \eta^2{}_\mathrm{G} = 0.030$), a significant medium effect was found for *Usefulness*, $F(2, 58) = 7.881, p < .001, \eta^2{}_\mathrm{G} = 0.136$). Post hoc tests revealed significantly better *Usefulness* ratings of the AR overlays compared to the Baseline ($t = 3.806, p_\mathrm{holm} = .001$) with a medium-sized effect of $Cohen's\ d = 0.695$ and compared to variant 2 (status bar with counts; $t = 2.882, p_\mathrm{holm} = .011$) with a medium-sized effect of $Cohen's\ d = 0.526$. Regarding the *Enjoyment* scale of Chen's TAM adaption, all variants achieved high ratings (Figure 4.5) with no meaningful effect ($F(2, 58) = 0.925, p = .402$, $\eta^2{}_\mathrm{G} = 0.014$). Similarly, all variants achieved medium-high ratings for *Intention to Use* (Figure 4.5) without relevant differences, $F(2, 58) = 1.553$, $p = .225$, $\eta^2{}_\mathrm{G} = 0.020$.

UX

With regard to *pragmatic UX quality*, all three variants received above middle ratings (Figure 4.5) with no significant differences between them, $F(2, 58) = 1.590, p = .213, \eta^2{}_\mathrm{G} = 0.030$. For *hedonic UX quality*, larger deviations ranging from above middle ratings (AR overlays) to medium-low ratings (baseline, status bar; Figure 4.5) with a significant large effect were found, $F(2, 58) = 10.447$, $p = .001$, $\eta^2{}_\mathrm{G} = 0.169$. Post hoc tests show significant higher hedonic quality with the AR overlays compared to the baseline ($t = 4.334, p_\mathrm{holm} < .001$) with a medium-sized effect of $Cohen's\ d = 0.791$ and compared to variant 2 ($t = 3.136, p_\mathrm{holm} = .005$) with a medium-sized effect of $Cohen's\ d = 0.572$.

Trust

The HMI variants received medium to medium-high assessments regarding *Understandability/Predictability* (Figure 4.6) with significant differences between the variants showing a medium-sized effect, $F(2, 58) = 8.128, p < .001, \eta^2{}_\mathrm{G} = 0.108$. Post hoc tests revealed significantly better *Understandability/Predictability* of the variant with AR overlays compared to the Baseline variant ($t = 3.810$, $p_\mathrm{holm} < .001$) with a medium-sized effect of $Cohen's\ d = 0.696$ and compared to variant 2 ($t = 3.048$, $p_\mathrm{holm} = .007$) with a medium-sized effect of $Cohen's\ d = 0.556$. Regarding *Reliability/Competency*, the variants obtained above middle ratings (Figure 4.6) without meaningful differences, $F(2, 58) = 2.309$, $p = .108$, $\eta^2{}_{\mathrm{G}} = 0.025$. Similarly, all three variants received medium, above middle ratings for overall *Trust in Automation* without a significant effect, $F(2, 58) = 1.803$, $p = .174$, $\eta^2{}_{\mathrm{G}} = 0.019$.

Risk and Safety

Risk was rated low throughout all variants (Figure 4.6) without meaningful differences, $\chi^2 = 1.869$, $p = .393$, $n = 30$. In accordance with that, the *Safety* scale received medium-high ratings in all conditions (Figure 4.6) without significant differences, $F(2, 58) = 0.677$, $p = .512$, $\eta^2{}_{\mathrm{G}} = 0.009$. The low risk values and the feeling of safety was often related to trust in the general capabilities of the automated system (e.g., P10: *"I am convinced of the capabilities of the system"*, P21: *"I trust the system. Unforeseen events were handled without problems."*) as well as the driving style (e.g., P16: *"[it] drives like me – safe"*; P24: *"The vehicle reacted with restraint in unusual situations. That was good."*; P28: *"Very relaxed way of driving [... and] good response of the vehicle to all situations."*).

Wellbeing and Nausea

With regard to *Wellbeing*, the single-item scale revealed positive assessments with medium-high ratings for the three variants (Figure 4.6). While no significant difference between the variants was found ($\chi^2 = 3.774$, $p = .152$, $n = 30$), descriptive statistics suggest that participants felt slightly better using variants 1 or 2 (Figure 4.6). This is similarly indicated by the *Nausea* scale (Figure 4.6). While only a few participants reported Nausea symptoms, there is a significant difference between the variants, $\chi^2 = 7.357$, $p = .025$, $n = 30$. Nausea symptoms occurred significantly more often with AR overlays compared to variant 2, $\textit{Conover T-Stat} = 2.838$, $p_{\mathrm{holm}} = .019$. However, the differences between AR overlays and baseline ($\textit{Conover T-Stat} = 1.845$, $p_{\mathrm{holm}} = .140$) and between baseline and variant 2 ($\textit{Conover T-Stat} = 0.993$, $p_{\mathrm{holm}} = .325$) are not statistically significant. Four participants (P19, P23, P25, P30) also described motion sickness symptoms verbally. While P23, P25, and P30 related the symptoms to generally watching at the digital screen during the ride, P19 accounted them particularly to the AR overlays: *"I think if I drove here longer, I might feel a little dizzy [...] from the color fields."*. P30 added that wearing an FFP2 face mask during the ride further influenced the occurrence of the symptoms.

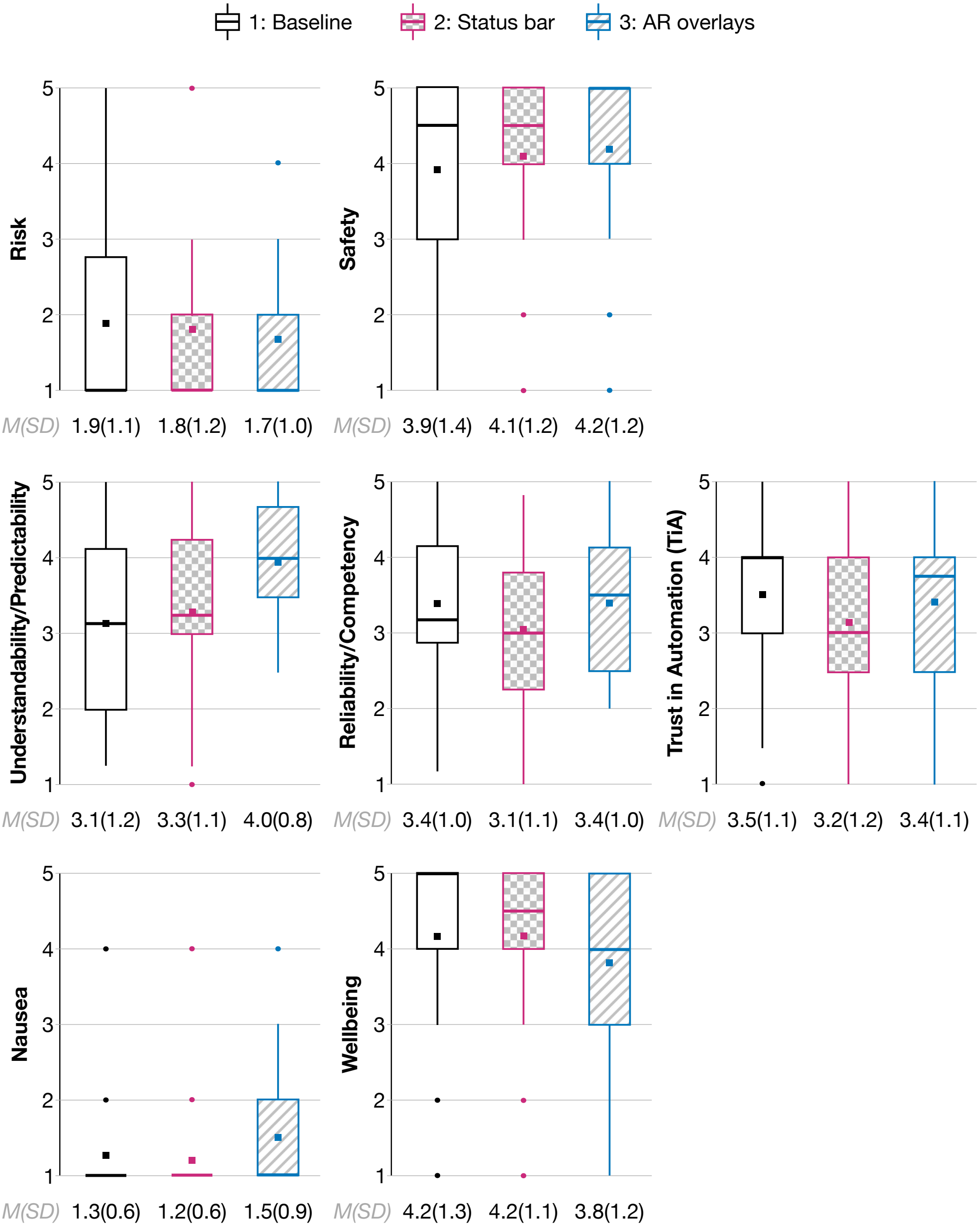


Figure 4.6 – Boxplots and $M(SD)$ of trust, risk, safety, wellbeing, and nausea scales for the three concept variants.

4.2.2 Qualitative Variant Assessment

Overall, variant 3 with the AR overlays (Figure 4.3: 3) was preferred by half of the sample. However, ten participants put the AR overlays on the last rank and ten rated variant 1 (baseline without information on detected objects; Figure 4.3: 1) as their favorite. Only five participants preferred variant 2 with the object counts in the status bar (Figure 4.3: 2). The following sections provide a detailed overview of the received qualitative feedback per variant.

Baseline (Variant 1)

Seven of the ten participants that opted for variant 1 (Figure 4.3: 1) found the visualization of surrounding objects generally unnecessary (e.g., P12: *"because if you don't drive yourself anyway, then it doesn't really need to display anything."*). P10 considered only the general system feedback ("System: OK") relevant and the object visualizations as a *"gimmick [. . .] unless it really has the consequence to intervene"*. Two participants argued that less information is better when it comes to trust in the technology (e.g., P28: *"the system seems more trustworthy even though there is less information available*").

Status Bar (Variant 2)

Twenty participants considered the status bar (Figure 4.3: 2) unnecessary (e.g., P23: *"I found this nice, but somehow just not helpful."*) – in contrast to five participants who described the count display as helpful. While some mentioned that the object counts increased perceived safety ($n = 3$) and trust ($n = 3$; e.g., P28: *"it's reassuring"*), others said it decreased both perceived safety ($n = 5$) and trust ($n = 3$). Since the counts jumped fast in some situations, eight participants perceived the fast refresh rate as irritating (e.g., P17: *"the display has made one restless [. . .] one is tempted to control the display"*). Without the matching overlays of variant 3, the meaning of the count display was unclear to 11 participants and left some with open questions and the desire for better contextualization of the information, e.g., P13: *"I would like to know how the car puts what it recognizes into context with the driving context"*.

AR Overlays (Variant 3)

The AR overlays of variant 3 (Figure 4.3: 3) were considered to be helpful for (better) understanding the driving situation ($n = 11$; e.g., P23: *"You could see at a glance what* *was happening and classify it much better"*) and to build trust in the system ($n = 12$; e.g., P13: *"One could better understand the complexity of the system. Therefore, more trust"*). However, in 16 of the 30 sessions, participants described the AR overlays to be either annoying, irritating, or distracting (e.g., P17: *"You can't enjoy the ride"*; P21: *"Too many colored boxes"*; P26: *"somewhat annoying display",*). In some sessions, participants reported that the AR overlays decreased perceived safety ($n = 6$) and their trust in the system ($n = 3$). This ambivalence was further observable in the interviews, where many participants weighed the variants' pros and cons.

4.2.3 Visualization Design

Twelve participants desired to have only objects relevant to the current driving scene visualized. Regarding the visual design of the overlays and object counts, participants considered distinct colors for object classes to be useful ($n = 24$) as well as various visualizations for critical objects ($n = 20$; e.g., P18: *"if a pedestrian would run in, [the overlay] becomes red for example"*). On the other hand, eight participants would have generally preferred fewer colors.

Amount and Type of Object Visualization

Asking participants which objects they would want to be visualized, only four voted for all objects. Eleven participants wanted only objects marked that have an impact on the vehicle's ride, and 12 preferred only marking hazardous objects (e.g., P7: *"[would be] clearer"*; P26: *"All objects is too much and too confusing"*). While two participants remained undecided, only P19 opted for no object detection visualization at all (*"too distracting"*). The latter is particularly interesting considering that in the general assessment, nine more participants ranked variant 1 with no object visualization as their favorite. With regard to the question which kind of objects should be displayed, participants mentioned vulnerable road users to be most important (bicyclists: $n = 23$, pedestrians: $n = 24$). Vulnerable road users are considered more important than general obstacles ($n = 20$), own driving trajectory ($n = 20$), other vehicles ($n = 19$), traffic lights ($n = 17$), infrastructure ($n = 12$), traffic signs ($n = 10$), or street markings ($n = 6$). Most participants ($n = 16$) preferred visualizing detected objects according to their hazard level. P13 suggested overlays with a transparency level according to their relevance or criticality. In contrast, ten participants considered a special visualization for hazards unnecessary. Two participants pointed out that objects that are not visualized can be regarded as unrecognized and, therefore, be a safety risk. P14 and P24 asked whether the system could detect animals (e.g., dogs).

Configurability

In general, the sample majority ($n = 23$) argued for configurable display settings allowing passengers to choose what, when, and how information is displayed. E.g., P13: *"I think when you probably use the system more often [… the display] could be a distraction and I might want to turn it off"*; P30: *"I think it's good if everyone could decide for themselves"*). P5 suggested that the visualization should turn itself off automatically after a specific time but can be turned on again by the passengers. Five participants proposed that the UI screens should be usable for other things, e.g., as a second screen for mobile work, information, or entertainment.

Additional Ideas

P1 would have liked to use the augmentations at night to enable a kind of night vision for passengers. P15 wanted an onboarding tutorial explaining the displayed information and functionalities to first-time users or on-demand. Similarly, P30 would have wanted a more detailed legend that explains, e.g., the meaning of the colors. P14 would have found it helpful if an indicator for the object's moving direction is displayed. P27 suggested acoustic warnings for critical situations so that passengers could prepare themselves, e.g., for occurring driving actions like emergency brakes. Many participants wished for improved (AR) visualization ($n = 6$) and suggested, e.g., not to display large overlays with "sharp" borders but rather, e.g., a spot or a dot (P28), a decent border (P4), a soft filling (P2), or a gradient or blur (P25), which they considered to be more convenient to look at and assumed to reduce flickering of the detection and consequent distractions.

4.2.4 Further Information Needs

A large part of the sample ($n = 14$) wanted to have information on the current location and the planned route, e.g., displayed on a map. Two participants suggested to have this on an extra display. Twelve participants would have liked to receive location-based information about their surroundings, such as descriptions of landmarks. Seven participants wanted driving-related data (e.g., current speed) since such information would increase their feeling of safety ($n = 5$). In contrast, three participants argued that they would need such information only at lower automation levels. P14, P17, and P20 would want the system to explain its (planned) driving actions (e.g., turning or parking). Several participants preferred controls for passenger interaction, e.g., touchscreen- or speech-based input options to customize the visualization display, navigate to a particular destination, or change the route or emergency buttons and functions to contact human support or a (remote) operator.

4.2.5 Wizard-of-Oz

After lifting the deception and explaining the WoOz setup, 22 of the 30 participants (73 %) stated that they believed that the vehicle was driving autonomously and that the driver was only there for safety reasons. An exploratory analysis revealed a significant correlation between participants doubting the WoOz illusion and their ATI scores (Spearman's rho: $r_\mathrm{s} = .411, p = .001$). This indicates that participants with a higher affinity for technology interaction were less likely to believe the deception. However, no other meaningful correlation was found between participants' belief in the autonomous ride and their quantitative assessment of the dependent variables. Thus, we do not differentiate the results based on that. In the following sections, we report detailed findings on the WoOz deception and cover story, participants' driving experience, environmental conditions of the test drives, and the prototype's fidelity.

Deception and Cover Story

Many participants who believed the deception commended the smooth, forward-looking, and defensive driving style (e.g., P2: *"The system mimics an exemplary driver"*; P25: *" When you drive yourself, it's usually not so smooth"*) and were surprised when we lifted the deception (e.g., P3: *"okay, I would have been sure that it drives automated"*). Some comments highlighted the importance of a thoughtful cover story. E.g., P27: *"It was good that you said [the AV] didn't have downtown approval yet, or I probably wouldn't have bought it off"*) and pointed out that the used vehicle's appearance and trust in a certain brand or manufacturer also affect the believability of system capabilities (P13: *"Such a new Mercedes ... that also helped. You tell yourself that it can do nothing wrong."*). However, others regarded the smooth driving style as an indicator that the vehicle could not have been driven by a machine only (e.g., P13: *"from my experience, that was too forward-looking"*). In some situations, that forward-looking driving style was not possible, or the driving wizard failed to conduct it. This led some participants to doubt the autonomous ride (e.g., P30: *"[the ride] was not anticipatory enough for me. So it was two times somehow that the traffic light was yellow and [the vehicle] decided to cancel at short notice"*). Other participants, who doubted the autonomous ride, missed visible sensor hardware indicating that the car is capable of autonomous driving or noticed the wizard's movements (e.g., P4: *"I heard [. . .] the use of the steering wheel when we were driving "*). P8 explained its doubts with prior knowledge of the current state of technology.

Ride Experience

At the end of the rides, 13 participants commended the positive driving experience (e.g., P12: *"Perfect. Not so abrupt [. . . but] nice and steady"*; P21: *"it was definitely a very pleasant ride [. . . and] very interesting"*). Nine participants felt safe because of the safety driver's presence (e.g., P3: *"I had confidence that the safety driver would intervene, if required."*). Four others said they felt safe because of the automated system only. Seven participants compared the ride in the (simulated) AV with being a passive passenger in a taxi or bus. However, some participants had different expectations (e.g., P14: *"I actually imagined autonomous driving to be [. . .] a softer way of driving"*). While a few participants felt unease due to the video see-through-based WoOz setup (section 4.2.1), others were not bothered by the setup at all (e.g., P22: *"I think that was totally realistic, [. . .] the image [. . .] was just fitting to the movements [. . .] it was [. . .] as if I was looking out of the front"*). Three participants mentioned that the view through the digital screen affected their perception of the ride (e.g., P11: *"You somehow feel it [...] as a faster ride on the screen than in real life"*).

Environmental Conditions

All test rides were performed during daytime in an urban area with moderate traffic density. Regarding the weather, most of the test rides were conducted under cloudy conditions ($n = 26$). In four sessions, it was rainy, in one snowy, and in 11 sessions, it was (partly) sunny. In the latter, six participants mentioned that the video feed was sometimes overexposed during the ride due to direct sunlight. During the rainy rides, the view out of the vehicle through the windows and consequently through the camera stream was (partly) impaired. However, as there was no heavy rain, the object detection kept functioning correctly.

Prototype Fidelity

Due to technical constraints (hardware and software), the prototype's performance was limited. Some objects were detected late or not at all (mentioned in 12 sessions). In such cases, it was not clear to participants how the visualized objects were selected (variant 2: $n = 12$; variant 3: $n = 9$). While we briefed participants that the tested HMI prototype's accuracy is limited due to its early development stage and unlinked from the (simulated) AV's sensors, several participants were disturbed by the offset of the vehicle's driving behavior and the visualization (e.g., P27: *"the car has already braked before the traffic light was even recognized* "). Participants also described some issues and limitations, e.g., occasional image stuttering of the video feed and visualization were considered irritating ($n = 6$; P28: *"It has quite a bit of flickering every now and then, which is a bit annoying."*;).

4.3 Discussion

The conducted on-road WoOz study provides design and prototyping learnings for suitable AV HMIs. Furthermore, it provides insights into the potential of transparent information and AV passenger experiences in general. In contrast to previous works, we investigated feedback from an AI-based system with an empirical study in a natural urban environment. The study featured the value of WoOz for context-based prototyping of HMIs that use real-time information and AR. We focus the following discussion on 1) object visualization and 2) passengers' information requirements regarding the tested HMI concepts, pursued by a discourse of 3) the WoOz-based prototype's potentials and considerations, as well as 4) limitations and future work.

4.3.1 Object Visualization Can Increase Acceptance and UX

With regard to Q4.1, qualitative and quantitative results confirm that visual system feedback on detected objects can increase AV acceptance and UX. Concerning the way this information should be presented in the AV windshield (Q4.2), most participants preferred the concept with AR overlays (variant 3) over both the baseline concept without information on detected objects (variant 1) and the status bar with object counts only (variant 2). The augmentations were considered helpful in understanding the context better and building trust in the system. This is confirmed by significantly higher understandability and predictability assessments. The results confirm the findings of the online study of Colley et al. [47], who reported increased situation awareness of drivers through (AR-based) visualizations in conditional automated driving. Significantly better evaluations of perceived usefulness and hedonic quality further support the potential of AR-based object visualization, exhibiting an improved UX. Furthermore, many participants reported that object visualizations increased the feeling of safety and trust in the autonomous system.

Although the general positive assessments of the AR overlays, they were in half of the conducted sessions described as too much, irritating, or distracting. This is in line with the findings of Kim et al. [134] on driver distractions induced by AR in vehicles with lower automation levels. The second largest group of study participants considered it sufficient to have only general information on the overall system state, which was the case in the baseline variant. Some participants would not want continuous system feedback at all since they would not want to be distracted from other tasks. The status bar with the object counts served for some participants as an explanation of the AR overlays (variant 3), but was considered not helpful when used alone (variant 2). Some participants did not want to have object visualization at all and favoured variant 1. A reason for this could be negative perceptions of the visualizations, especially when occurring errors degrade the attributed system capabilities (see Section 4.3.4), which is in line with the findings of another online study by Colley et al. [48]. Furthermore, the rather salient design of the bounding box overlays might have been perceived negatively. Less obtrusive designs (e.g., less salient colors or colored borders only) might be more suitable. Only a few participants would want an intermediate solution, e.g., in the form of the status bar offered by variant 2.

The results generally confirm the potential of transparent system feedback to increase AV acceptance and UX, but not all people would want this information (all the time). Furthermore, the assessed concepts are only early-stage variants. They cover only a small part of the vast possibilities, especially regarding higher-performing hardware and algorithms that might enable more advanced and accurate object visualizations.

4.3.2 Passengers Want Configurability and Travel Information

Most participants argued for configurable display settings providing passengers with options to select what, when, and how information on the environment and the AVs' reasoning is displayed (Q4.2). This aligns with the findings of Oliveira et al. [185], who pointed out that AV passengers might not want to have contextual information displayed permanently. Participants in our study argued for configurability and on-demand information retrieval (i.e., the HMI should allow them to turn certain information on and off). Regarding the "what", most participants preferred visualizing only objects with an impact on their ride and a visual classification according to their hazard level instead of having permanent visual feedback on detected objects. Future work may investigate respective visualization designs.

The hedonic and pragmatic UX assessment of the three concept variants is in comparison to the UEQ benchmark [222] relatively poor. Qualitative results indicate that this can be (partly) attributed to missing expected travel information, e.g., current location, planned route, and upcoming maneuvers. I.e., since the tested concepts solely focused on providing information on detected objects and system status, participants missed journey-related information. As this was mentioned by almost half of the sample, travel information seems crucial in AVs. This result can be linked to findings of related work on lower automation levels, e.g., [82]. We recommend future work to consider the interplay of novel (information and visualization) concepts with such expected information and to investigate them as a part of holistic interaction concepts.

4.3.3 Wizard-of-Oz Setup and HMI Prototype Can Serve as Mutual Enablers

The applied WoOz-based setup served two purposes. First, it paved the ground for creating an on-road AV simulation for context-based prototyping and evaluation. Second, it enabled using a computer-vision-based real-time object detection system as an HMI component of a futuristic AR windshield. Nevertheless, WoOz settings also come with methodological challenges [19, 175] that need to be considered, such as keeping up the deception throughout the study. Considering these and the lack of comparable benchmarks, we regard having 73 % of participants believe that the vehicle was driving autonomously until the end of their sessions as evidence of a successful application of the WoOz paradigm. Our exploratory analysis did not reveal statistically relevant correlations of the dependent variables with participants' belief in the WoOz illusion, which is why we did not differentiate the results based on that. Further associated aspects and limitations are discussed in Section 4.3.4.

In addition to a smooth, defensive, and proactive driving style as recommended by related work [11, 175], we found having suitable hardware with a modern and technologically-advanced appearance (i.e., a vehicle believed to be capable of autonomous driving) as quite supportive of keeping up the deception. Overall, a thoughtful cover story seems to be a crucial part of the WoOz deception. In our case, we consider shifting participants' attention toward the futuristic HMI prototype beneficial. To do this, we told them that we were evaluating new concepts for not yet available hardware components (the AR windshield) and were, thus, requiring the TV-based setup. As a result, WoOz and the windshield interface prototype were mutually beneficial and enabled their successful application.

Nonetheless, not all participants believed the story. Reasons for the doubts can be allocated to, e.g., difficulties in constantly maintaining the defined driving style (e.g., when unexpected events occur), previous knowledge of participants on the state of technology, or observations of participants (e.g., driving-related noises of the wizard). Furthermore, while many participants described the test rides as pleasant, some participants noticed that the video see-through based setup made them feel at unease. This might have been due to the indirect view out of the vehicle and the camera's offset, as well as to the display of the visualizations.

To sum up and answer Q4.3, the created WoOz-based prototyping framework served as a suitable basis for this study and may be used and adapted to address similar questions. We recommend future work to thoroughly craft their prototypes, setups, and cover stories and leverage their symbiosis. The prototyping approach can be optimized for future studies according to our descriptions and findings.

4.3.4 Limitations and Future Work

In the following sections, we discuss the limitations of this work and the consequent potential for future work regarding 1) the study sample, 2) the HMI prototype, and 3) the applied WoOz approach.

Study Sample

The study sample is characterized by a medium-high affinity for technology interaction. While this is considered a common phenomenon in HCI studies [94], it might impair external validity and affect the belief in the WoOz deception, as the correlation revealed by our exploratory analysis suggests. Furthermore, as is often the case in usability testings, study participants experienced the evaluated system and HMI concepts for the first and only time. However, users' attitudes toward certain aspects can change over time. Future work might conduct long(er)-term studies to account for this circumstance. We also want to note that the study was conducted during the COVID-19 pandemic. Therefore, we applied precautions and hygiene measures (e.g., distancing, wearing medical/FFP2 masks, disinfection of surfaces and hands) and followed local and national authorities' regulations and recommendations. While we consider the pandemic's actual effect on the study results to be minor, it might have affected the sample composition as, e.g., only people without fear of COVID-19 might have signed up for the study in the first place.

HMI Prototype

The evaluated concept variants had a relatively narrow focus on object detection visualization on a (prototyped) windshield interface. We assume that this affected the overall assessment of the rides and visualization concepts. Since the targeted acceptance and UX challenges cannot be addressed in this narrow scope alone, future work should consider the integration with "holistic" HMI concepts (e.g., including visual and auditory passenger information on the planned route and upcoming stops; Chapters 3 [85], 6 [87]). This would also allow for further investigation of the design space, e.g., in terms of other visualization concepts such as 3D representation in a GUI-based map [246, 264], situation prediction visualization [48], and other feedback modalities such as auditory, kinesthetic, or tactile [126]. Future work may also investigate the HMI configurability suggested by participants and identify relevant situations, maneuvers, objects, or levels of criticality in which information and explanations would be (not) beneficial. As mentioned in Section 4.1.1, the prototype's hardware and performance were limited and consequently not as powerful as cutting-edge sensing systems. This resulted, to some extent, in flickering, missed objects, and classification errors. While the algorithm proved quite robust on rainy rides, extreme lighting conditions (e.g., direct sunlight) resulted in overexposure of the video feed and consequent impairments of sight and object detection. Nevertheless, the used hardware and algorithms served as a suitable basis for the early concept study, the straightforward realization of the AR windshield prototype with real-time information visualization, and the initial investigation of our focus questions in an early development phase. Future work may use more powerful industrial hardware and software along with more graphical and computational processing power to enable the use of larger and higher-performing models. Furthermore, adding additional object classes to the model (e.g., animals, construction sites, or hazardous objects) and investigating other visualization approaches might be interesting depending on examined scenarios and conceptual considerations.

Wizard-of-Oz Approach

Since actual AVs are still only available under limited conditions in urban environments, we applied the WoOz paradigm to create a prototyping framework that enabled us to consider the dynamic urban context in our investigation. The approach offers several advantages – especially concerning the evaluated real-time visualization prototype. However, it also poses challenges regarding objectivity, validity, and reliability [175]. While we aimed to control the study as much as possible, particularly dynamic influences cannot be ruled out completely. Furthermore, some of our participants were not fully convinced by the WoOz deception. Since we found no statistically relevant correlation between participants' belief in the WoOz deception with our dependent variables, we did neither exclude data nor create groups based on this. However, we cannot rule out possible effects on the results. Besides the challenges and limitations mentioned, future work may use the described WoOz approach and AR windshield prototype with the reported learnings to conduct further empirical studies. Such studies could investigate AV passenger experiences in a real-world context and HMI concepts relying on real-time information, e.g., visualizations of scene detection, scene prediction, and maneuver planning [48]. Considering the effort to conduct a WoOz study, researchers might, in a first step, formatively evaluate their designs with simpler study designs (e.g., online or simulator studies). In our case, for example, the desire for configurability could have been discovered earlier so that the results could have been incorporated into the subsequent WoOz study. The framework may also be used to prototype AR-based infotainment systems that provide contextual information, e.g., on landmarks or other points of interest [22]. Furthermore, future work could focus on the method itself and investigate the effect of participants' belief in the WoOz deception, e.g., by comparing one group that is told frankly about the system's actual capabilities with another group that is told the WoOz cover story.

4.4 Conclusions

Suitable HMI concepts are required to address AVs' acceptance and UX challenges. The conducted on-road WoOz study with 30 participants evaluating early visualization concepts for a windshield interface confirms the potential of transparent communication and object detection visualization to increase the acceptance and UX of AVs.

System feedback on detected objects was deemed useful, and AR-based visualization, in particular, significantly increased the system's understandability and predictability, perceived usefulness, and hedonic quality. However, in line with related work from online surveys, lab studies, and other automation levels, we found that (permanent) system feedback can also annoy, irritate, or distract passengers. We identified making the information configurable for individual user requirements and accessible on-demand as a promising approach to address this challenge. In addition, as travel-related information (e.g., current location, planned route, and upcoming stops) is essential in driverless vehicles, it needs to be investigated how transparent system feedback can be integrated with such information into holistic AV HMI concepts.

The applied video-based WoOz approach provides a suitable framework for prototyping both AVs and (AR-based) windshield interfaces with real-time information visualization. However, it poses technological and methodological challenges. A compelling cover story is essential for keeping up the WoOz deception and the study's success. It can be supported by fitting hardware (e.g., a modern vehicle) and an appropriate "AV-like" driving style.

To sum up, this work contributes to the human-centered design of human-AV interactions. It demonstrates a straightforward WoOz-based method for context-based prototyping of (AR-based) real-time AV HMIs that is suggested to become adopted and advanced by future work. Furthermore, it provides learnings and practical recommendations for system design and future studies.

4.5 Chapter Summary

This chapter demonstrated the prototype and WoOz-based on-road evaluation of a futuristic windshield HMI concept that visualizes real-time object detections via AR. In a mixed-methods within-subjects study ($N = 30$), participants assessed three early-stage concept variants of the windshield HMI prototype to explore whether object detection visualization can counteract acceptance challenges.

Inspired by the setups of Karjanto et al. [131] and Detjen et al. [64], we created a straightforward WoOz setup in a minivan (Mercedes-Benz EQV). As we used a rental car, a decisive requirement for us was – in contrast to related work – that we could quickly deconstruct the setup without leaving any traces of the modifications. The setup separated the front seats from the back area with a black curtain mounted at the vehicle's ceiling and a 43" TV mounted at the front seats' headrests. As a result, passengers in the back could neither see the windshield nor the (co-)driver. The TV displayed a video stream of a webcam mounted in the windshield, enabling a video-see-through view out of the front.

The TV-based setup enabled us to prototype a futuristic AR windshield HMI which we used to display information on the system's capabilities. In particular, the HMI provided information on computer-vision-based object detections with three variations: 1) no information (baseline), 2) abstract status bar with counts of the detected object classes, and 3) AR overlays. The information was displayed in real-time on the TV and derived by an embedded AI system with object detection algorithms performed on an Nvidia Jetson Nano. The resulting study design enabled us to craft a cover story that shifted participants' focus on the AR windshield prototype and away from the WoOz setup. We regard this as a significant contributing factor to the successful application of the WoOz method. In the presented study, almost three out of four participants believed the deception of an autonomous ride.

The study results confirm that transparent system feedback, particularly AR-based visualizations, can increase understandability, perceived usefulness, and hedonic UX. However, the amount and the timing of the provided information are crucial. Participants discussed the pros and cons of the information and demanded options for customization since the (permanent) information display might also be too much, irritating, or distracting. Considering other information demands that AV users have during the ride with an AV, we deem this specifically relevant. For example, many participants missed 'classic' information such as current location, estimated arrival times, or upcoming stops, which led us to conclude that future work needs to investigate such HMI concepts from a holistic perspective on human-AV interaction.

This chapter offers two main contributions to **RQ1**. First, it confirmed that information transparency (e.g., facilitated through AR-based object detection visualization) can increase AV passengers' acceptance and UX. However, the results also confirm findings of related work (e.g., [185, 197]) that the amount and timing of information displayed are elemental. Second, it provided insights and recommendations for the design of in-vehicle HMI concepts for future AVs. For instance, we recommend considering AV passengers' 'classic' travel information demands when designing new concepts and options for adjusting system feedback. Contributing to **RQ2**, this chapter demonstrated a straightforward approach to creating a flexible and easily deconstructable WoOz setup that can be used for context-based interface prototyping and evaluating real-time object detection systems. Based on its documentation, the approach can be used and adapted by other researchers and serve as a framework for further studies. Furthermore, we reflected on the challenges of the WoOz method and provided learnings and recommendations on how to overcome them.

5

Graphical vs. Conversational Interfaces for On-Demand Rides

Following the goal of counteracting AMoD acceptance hurdles (RQ1), most previous work focused on general acceptance aspects of AVs (e.g., [24, 41, 181]). Only few studies investigate the actual interaction, i.e., the design and evaluation of HMIs (e.g., [69, 135]). However, at the latest when the technology is ready for market introduction, the HCI community needs to be able to provide future practitioners, manufacturers and service providers with guidelines and recommendations for human-centered design and development.

In this chapter, we explore concrete approaches for HMI design and evaluation to tackle potential acceptance and UX hurdles of AMoD systems from an early

THE CONTENTS OF THIS CHAPTER WERE PUBLISHED PREVIOUSLY AS PART OF:

[86] **Lukas A. Flohr**, Sofie Kalinke, Antonio Krüger, and Dieter P. Wallach. 2021. Chat or Tap? – Comparing Chatbots with 'Classic' Graphical User Interfaces for Mobile Interaction with Autonomous Mobility-on-Demand Systems. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. `https://doi.org/10.1145/3447526.3472036`

development stage. As we are considering the complete user journey, we focus on mobile applications capable of accompanying users through all usage situations and confidentially providing individualized information – also during shared AMoD rides. After registration, user journeys in AMoD systems can be structured in three phases along time. For comprehensive support, the UI should assist users in all of these phases: (1) before the ride: the UI provides functionalities for information and booking; (2) during ride: the UI serves as a travel companion; and (3) after the ride: the UI assists with offboarding, rating the service etc. To create optimal interactions, it is necessary to be able to evaluate a UI along the whole journey.

Currently, 'classic' GUIs with touchscreen-based interaction can be considered as the status quo for interacting with mobility applications on personal devices (e.g., Figures 2.13:i, vii). However, AI-infused systems, comprising assistive dialog-based interaction concepts also promise to provide a good or even better fit when it comes to acceptance of new technology. Such support can be accessed via conversational UIs (CUIs) like chatbots (text-based) and virtual assistants (speech-based). CUIs enable people, similar to talking or chatting with a real person, "to interact with smart devices using spoken language in a natural way" [164, p. 1].

The question arises whether CUIs are as good as or even better suited than 'classic' GUIs for interacting with AMoD systems on mobile devices. Especially when it comes to establishing a sensitive communication between users and systems in new usage situations, CUIs seem to provide a promising approach (e.g., [255]). Depending on the situation, the two approaches might also be combined meaningfully. For instance, in many modern cars, drivers can either tap on a touchscreen or talk to a virtual assistant while driving. However, considering shared rides (i.e., multi-user scenarios) and users' privacy requirements (e.g., [30]), speech-based interaction with virtual assistants does not seem to be a good basis for interacting with AMoD systems on personal devices. Instead, classic GUIs and chatbots appear favorable. While GUIs and chatbots are intensively investigated across domains, literature lacks comparisons of the two approaches. If available at all, the transferability of existing comparisons (e.g., [255]) to the AMoD domain is limited. In the context of our general reserach question, **RQ1**, this chapter aims to fill the identified research gap and, consequently, investigates the following question.

Q5.1 Are 'classic' GUIs or chatbots better suited to provide the basis for mobile human-AMoD interaction in terms of UX and user acceptance?

This chapter's contribution is twofold. We illustrate and compare the strengths and weaknesses of chatbots and 'classic' GUIs – in general and specifically for the AMoD domain – based on existing literature, expert evaluations, and a comparative user study. Second, we demonstrate how interface prototypes for AMoD systems can be evaluated and holistically compared in early development phases while considering the complete user journey in a video-based simulation environment.

5.1 Material and Method

This work is part of a larger research project on AMoD and future mobility concepts (project APEROL). Within the project, an iterative human-centered design process [125] is applied. User research activities conducted within the projects – comprising, e.g., citizen dialogues ($N = 76$), large-scale online questionnaire studies ($N_{\mathrm{I}} = 456$, $N_{\mathrm{II}} = 148$) and stakeholder interviews – are combined with literature research and (technical) requirements engineering activities to form the basis for the design of a fictitious AMoD service. Resulting artefacts like personas, user journeys, and test scenarios form the foundation for the development of various UI concepts (e.g., mobile apps and in-vehicle UIs) for interacting with a future AMoD service in a confidential and private manner.

The work presented in this chapter concerns the design and comparison of a GUI-based and a chatbot-based AMoD companion app. Although the two interaction concepts can be considered to be rather different, they both offer suitable approaches to interact with AMoD systems using mobile devices (Section 2.1.2). To ensure a fair comparison of the two higher-level concepts, equivalent and representative prototypes are designed and, in a first step, evaluated in separate expert studies and then improved based on the findings. In a second step, the iterated prototypes are compared in a simulator user study.

5.1.1 Design Process and Prototypes

Below, we provide details about the GUI and chatbot prototypes' design process. Both prototypes' design was optimized for an iPhone X, which was used as a test device in all three studies reported in this chapter. See Figure 5.1 for an overview of the design process and Figure 5.2 for screenshots of the final prototype versions used in the user study.

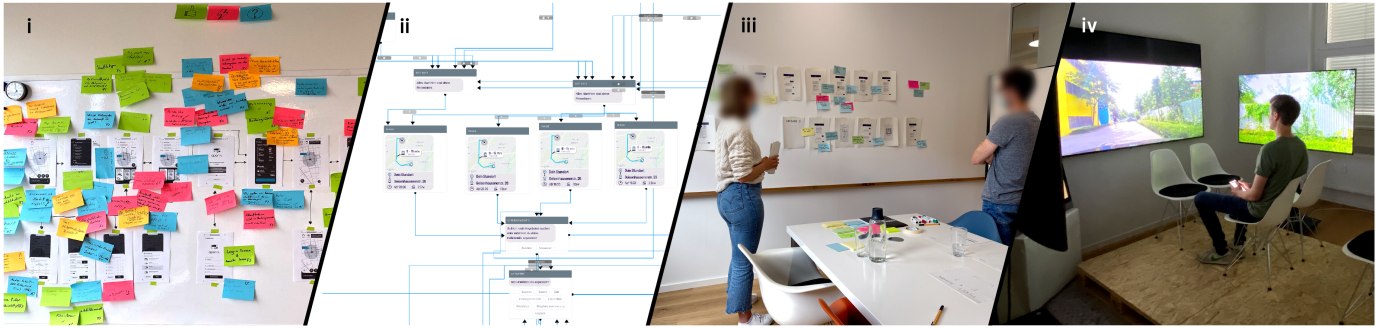


Figure 5.1 – Selected phases of the design process: initial conceptual designs were discussed in stakeholder workshops (i), iterated, and then used to create high-fidelity prototypes using Sketch and Botmock (ii). The GUI and chatbot prototypes were evaluated in expert studies (iii), improved, and then compared in a between-subjects user study with an immersive video-based AV simulator (iv).

GUI

At first, the GUI prototype was developed based on the mentioned preceding user research activities. After a first ideation phase, we created conceptual wireframes for an app that supports future AMoD users along the whole user journey. We used the wireframes to collect initial feedback from designers, developers, and other stakeholders (public transport providers, logistics service providers, city councils, and urban planners) in concept workshops (Figure 5.1:i). Subsequently, the concept was iterated, and a high-fidelity prototype was created using Sketch v.67.

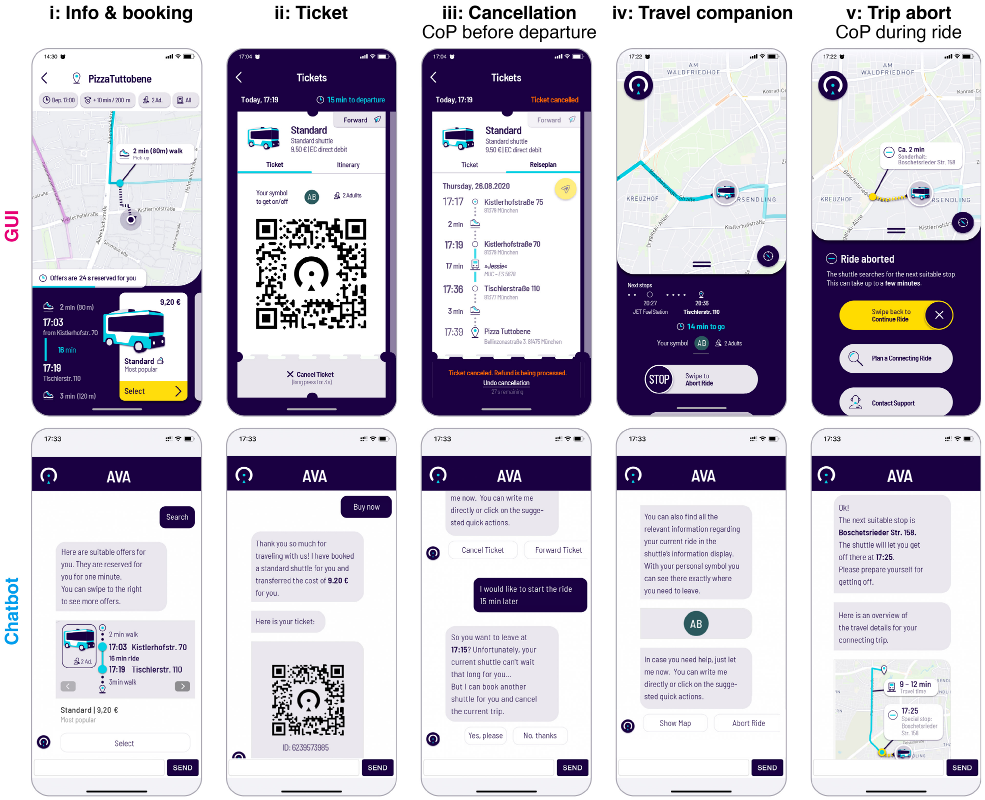
The prototype featured a map-based main view to entering travel details and requirements (e.g., departure time, destination, number of travelers, shuttle class, temporal and local flexibility). The service was designed to offer users three options for rides based on their input. The options differed, e.g., in terms of departure and travel times, shuttle classes (e.g., standard shared or express), prices, and walking times. Furthermore, the app featured a ticket wallet and an in-build navigation functionality (e.g., to find the pick-up point or the final target destination). During the ride, the GUI was designed to provide real-time travel information (e.g., current location, estimated arrival time), access to support and emergency functions, and the option to abort the current ride at the next possibility.

After optimizing the 'classic' GUI based on the findings from the expert study (Figure 5.1:iii, Section 5.2.1), it served as a foundation for the chatbot prototype.

Chatbot

For the design of the chatbot's personality and conversation flow, existing guidelines and recommendations [103, 107, 232, 245] were considered. Besides that, an additional online-survey on CUIs for AMoD ($N = 70$) was conducted to gain insights about potential users' preferences.

Based on the survey results, the chatbot's tone of voice (queried using the four primary tone of voice dimensions of Moran [172]) was defined to be rather respectful than irreverent but balanced in terms of funny/serious, formal/casual, and enthusiastic/matter-of-fact. Regarding the bot's visual appearance, participants preferred the shuttle service provider's logo over a human, robotic or abstract avatar. Most participants preferred the chatbot to take the service's perspective when communicating (e.g., "your shuttle will be there in 5 minutes") rather than the shuttle's perspective (e.g., "I'll be there in 5 minutes"). Despite using natural language to communicate with the



Here are suitable offers for you. They are reserved for you for one minute. You can swipe to the right to see more offers.

Thank you so much for traveling with us! I have booked a standard shuttle for you and transferred the cost of **9.20 €** for you.

Here is your ticket:

ID: 6239573985

me now. You can write me directly or click on the suggested quick actions.

I would like to start the ride 15 min later

So you want to leave at **17:15**? Unfortunately, your current shuttle can't wait that long for you... But I can book another shuttle for you and cancel the current trip.

You can also find all the relevant information regarding your current ride in the shuttle's information display. With your personal symbol you can see there exactly where you need to leave.

In case you need help, just let me now. You can write me directly or click on the suggested quick actions.

Ok! The next suitable stop is **Boschetsrieder Str. 158.** The shuttle will let you get off there at **17:25**. Please prepare yourself for getting off.

Here is an overview of the travel details for your connecting trip.

The shuttle searches for the next suitable stop. This can take up to a few minutes.

Figure 5.2 – Selected states of the final prototypes used in the simulator user study (translated from German to English; UI design and illustrations: ©Ergosign; maps: ©Mapbox, ©OpenStreetMap). Top: 'classic' GUI; bottom: chatbot (text-based CUI).

chatbot, participants liked the bot to provide supporting graphics and quick action buttons as shortcuts.

The resulting chatbot prototype 'AVA' (Autonomous Vehicle Assistant) was created with Botmock Conversation Designer. Like the GUI, the chatbot was also evaluated with a mixed-method expert study (Figure 5.1:iii) and optimized based on the results. In a second step, we called in an immersive video-based AV simulator (Figure 5.1:iv; [85]) to test and compare both concept prototypes holistically considering the complete user journey – from planning and booking over experiencing the booked ride in the simulator to reaching the target destination and rating the service quality (Figure 5.3). All studies were conducted in German.

5.1.2 Expert Studies

Formative evaluations with HCI experts provided the basis for subsequent prototype iterations for each concept. The experts were recruited from internal design teams that were external to the project. I.e., participants did neither take part in creating the prototypes nor were familiar with the designs before participating in the study.

Participants

Six participants (5 female, 1 male, 0 diverse, 0 n/a) between 24 and 33 ($M = 28$, $SD = 3.8$) and an average working experience of five years took part in the GUI expert evaluation. The Affinity for Technology Interaction (ATI) [94] indicated rather high technology affinity among the sample with $M = 4.3$ ($SD = 0.5; 1 = min; 6 = max$).

In the chatbot's expert evaluation, five participants (3 female, 2 male, 0 diverse, 0 n/a) with an average age of 28 years ($min : 24; max : 32$) and an average working experience of about four years took part. Again, the ATI [94] indicated high technology affinity among the experts with $M = 4.7(SD = 0.5; 0 = min; 6 = max)$.

Procedure

Each expert participated in a single-session with an experimenter and a note taker. The sessions were structured in four phases: (1) pre-session interview (demographics, experience), (2) scenario- and task-based usability test with thinking-aloud and live note-taking on a whiteboard, (3) User Experience Questionnaire (UEQ; 26 bipolar items; seven-point scale; scales: Attractiveness, Stimulation, Novelty, Perspicuity, Efficiency, and Dependability) [150, 220], (4) semi-structured interview and discussion of notes (Figure 5.1: iii).

Following a mixed-method approach [52], we collected both quantitative UX metrics using the UEQ and qualitative aspects by conducting a semi-structured interview and discussion at the end of the sessions to identify optimization potential. Based on the feedback and consecutive design iterations, we wanted to assure a certain quality level of the prototypes for their fair comparison in the subsequent user study.

5.1.3 User Study

For the comparative evaluation of the two concepts, we considered context as a crucial component to achieve valid assessments. As stated above, there are several approaches to simulate the not yet available context of using AMoD services. Since we wanted to create a reproducible experience that lets participants encounter a complete user journey in an AMoD system, we decided to go for a lab-based simulation. We adapted and enhanced the video-based method described by Flohr et al. [85] to create a straightforward mock-up of a shared AV (Figure 5.1: iv; Figure 5.3).

Experimental Design

We presented a typical use case of future AMoD systems in the user study: booking and taking a single trip in a shared AV. Here, we put special attention to two potential scenarios. (1) Journeys with 'happy paths', i.e., journeys without exceptional or error conditions and no occurring changes of plans and – since increased flexibility would be a primary reason for users to switch to AMoD [192] – (2) journeys with user-initiated changes of plans (CoP; changing the destination and departure time of an already booked trip with a shared shuttle).

Based on the insights gained from related work and our expert studies, we assumed that scenarios with 'happy paths' can be conducted without major effort or frustration using the GUI. However, we conjectured that in CoP scenarios, the strength of the chatbot's dialog-based conversation style would come into play. Here, the bot could provide a dialog-based explanation for the possible restriction (in shared rides, desired changes can only be made with consideration for the other passengers) and guide the user to a solution. Based on these assumptions, we derived the following hypotheses regarding the user acceptance of the AMoD system:

H1.1 Acceptance is higher when using the chatbot than when using the GUI in CoP scenarios.

H1.2 Acceptance is higher when using the GUI than when using the chatbot in 'happy path' scenarios.

As shown in other contexts [2, 42], UX is linked to user acceptance. That is why we assumed an effect analogously in UX metrics. We further assumed that different ratings in the test conditions should merely be based on task-oriented interaction quality [222] and, therefore, focused on pragmatic quality to derive the following hypotheses:

H2.1 Pragmatic quality of the chatbot is higher than that of the GUI in CoP scenarios.

H2.2 Pragmatic quality of the GUI is higher than that of the chatbot in 'happy path' scenarios.

To assess the hypotheses, the study used a counterbalanced mixed design with a between-subjects factor of the UI concept (GUI vs. chatbot) and a within-subjects factor of the scenario ('happy path' vs. CoP).

Data Collection

Quantitative and qualitative data was measured using a digital questionnaire after each ride. Acceptance was assessed using the questionnaire by Van der Laan et al. [256] (9 bipolar items with 5-stage scale; scales: Usefulness and Satisfying) as well as with the Intention to Use scale of Chen's [41] adaption of the Technology Acceptance Model (TAM) to the AV domain (2 items; five-point Likert-type scale). Pragmatic Quality was assessed by the respective subscales of the UEQ [150]: Perspicuity, Efficiency, and Dependability.

Another goal of the study was targeted at an exploratory evaluation. Since the UI would be the main touchpoint between the AMoD service and the users during the whole journey, we were curious how the UI concept might also affect hedonic aspects, users' emotional constitution, and their trust in the system. Therefore, we also evaluated the UEQ [150] scales Stimulation, Novelty, and Attrativeness as well as users' Trust and Emotion after each ride. The applied Trust scale (3 items; five-point Likert-type scale; [171]) is also part of Chen's [41] adaption of the TAM. Emotion was assessed by emotion curves drawn by participants on an adapted version of the template used by [135].

Furthermore, we intended to assess the chosen methodological approach and its suitability for evaluating AMoD systems. Therefore, we collected participants' presence perception using the Igroup Presence Questionnaire (IPQ; 14 items; seven-point Likert-type scale; [224, 225]) and a single item (five-point Likert-type scale; *'I found the ride in the simulator realistic.'* [85]) as well as participant's Wellbeing (single item with five-point Likert-type scale; *'I felt comfortable during the ride.'* [85]), and the occurence of Simulator Sickness (single item with five-point Likert-type scale: *'I felt nauseous during the ride in the simulator.'* [101]). A semi-structured interview on the AMoD service, the UI concepts, and the simulation closed the evaluation session.

Participants

To achieve sufficient power $\geq .80$ with an alpha error rate of $\alpha \leq .05$ for calculating inferential statistics (planned contrasts and ANOVAs with repeated measures and within-between interactions), a required sample size of 34 participants was determined a-priori using G*Power 3.1 for Mac. Considering practical impacts, medium effects according to Cohen [46] were assumed.

In total, 35 participants took part in the study. However, one participant's data had to be excluded from the analysis due to prototype and simulation errors during the session. This resulted in a sample of $N = 34$ participants (15 female, 18 male, 0 diverse, 1 n/a) between the ages of 19 and 61 ($M = 29.9$, $SD = 9.8$). The ATI short scale (ATI-S) [268] indicated rather high technology affinity among the sample with $M = 4.2$ ($SD = 1.3; 0 = min; 6 = max$). All participants were recruited via online postings and received financial compensation (30 €).

Procedure

Prior to the study, participants were randomly (counterbalanced) assigned to one of the between-subjects groups leading to 17 participants testing the chatbot and 17 participants testing the GUI. At the beginning of the study session, participants received a detailed briefing on the study's purpose and procedure and the potential side effects of simulator sickness. Afterwards, they signed an informative participation consent form and filled out a pre-questionnaire (demographics, ATI-S). Participants were encouraged to think aloud and ask questions whenever they wanted and – in case they felt at unease – to pause or quit the study at any time without consequences.

For the task-based testing, participants experienced two AMoD journeys in the lab-based setup. The order of the journeys was randomized (counterbalanced). For each of the journeys, they were provided with a contextual scenario. Their task was to travel with a friend to a public park (journey 1) and back home (journey 2) using the AMoD service. Therefore, they had to enter the respective target location and departure time and select one out of the three options provided by the app.

After booking, participants used the app's build-in navigation functionality, checked in at a paper-prototyped door UI using the received ticket QR code, and took the ride in the AV simulator. In the 'happy path' journey, participants did not receive explicit instructions on what to do during the ride. I.e., they were free to, e.g., monitor the information display or to just enjoy the (simulated) ride. In the CoP condition,

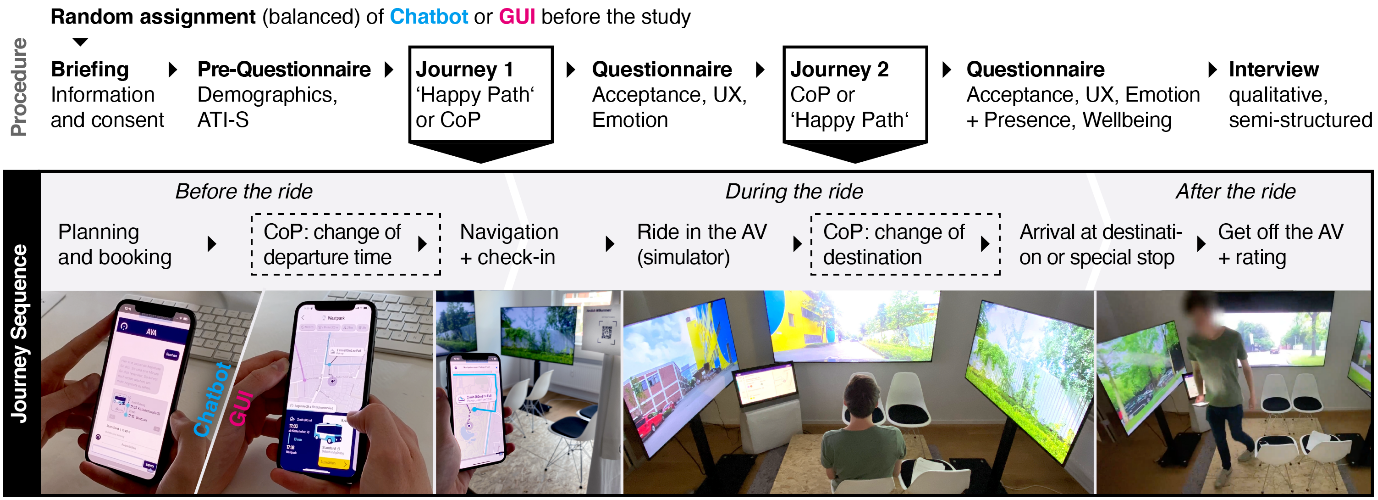


Figure 5.3 – Study procedure and journey sequence illustrating how participants experienced the prototypes in the immersive video-based lab setup. Scenario order ('Happy Path' and then CoP or vice versa) has been randomly assigned (counterbalanced).

two CoPs occurred. (1) A change of the departure time for an already booked ride implied canceling the initial booking prior to departure and booking a new ride. (2) A change of the target destination during the ride required participants to abort the ongoing ride and book a connecting ride to the new destination (a restaurant). When they reached the target destination ('happy path'; after a 14 min ride) respectively the special stop (CoP; after a 7 min ride), participants checked-out and rated the service using the app (Figure 5.3).

After each ride, participants assessed acceptance, UX, and trust using a digital questionnaire and reported their emotional changes during the journey by drawing an emotion curve. At the end of the session, participants evaluated presence perception, wellbeing, and simulator sickness with the digital questionnaire and took part in a semi-structured interview. Each session took about 75 to 90 minutes and was conducted by one experimenter and one note-taker.

Apparatus

For a context-based evaluation of the concepts, we adapted the immersive video-based driving simulator proposed in Chapter 3. In contrast to the setup of Chapter 3, large TV screens with 4K resolution instead of projectors formed the simulator's basis. The video footage was recorded with three action cameras (left, center, right) while driving through an urban environment and supplemented with audio footage (e.g., opening/-closing doors, signal sounds, voice prompts). To enhance the simulation's immersion, the screens were configured to form a closed space embracing an elevated platform with a seating group (Figure 5.1: iv; Figure 5.3). A 26-inch in-vehicle information display – illustrating a timetable, upcoming stops, and a map-view with the AV's current position and planned route – complemented the setup and the UI concepts.

5.2 Results

Created prototypes were evaluated in mixed-method expert studies ($n_{\text{GUI}} = 6$; $n_{\text{Chatbot}} = 5$) and improved based on the findings (i.e., discovered usability problems were solved and the experts' suggestions for improvement were implemented). In a second step, the iterated prototypes were compared in a between-subjects simulator study ($N = 34$).

5.2.1 Expert Studies

Both the GUI and the chatbot prototypes scored above average to excellent ratings in the UEQ subscales [150, 220] (scales range from -3 to 3). Qualitative findings revealed minor and cosmetic usability problems.

GUI

According to the UEQ Benchmark [222], the GUI received excellent ratings in terms of Attractiveness $(M(SD) = 2.1\ (0.7))$, Efficiency $(M(SD) = 2.0(0.7))$, and Dependability $(M(SD) = 1.8(0.4))$ as well as good ratings for Perspicuity $(M(SD) = 1.9(0.3))$, Stimulation $(M(SD) = 1.5(0.7))$, and Novelty$(M(SD) = 1.4(1.1))$ by the six experts.

The use of commonly established visual and conceptual patterns (e.g., form and navigation design, or auto-suggestions while entering start and target locations) was commended by three experts. Several non-standard functions (e.g., reservation time, trip abort) were, however, described by five experts to be ambiguous since the GUI lacked explanations.

One participant questioned whether specific functions (e.g., search for new rides) need to be always visible (e.g., during the ride) and recommended an adaptive behavior with respect to the context of use. Four experts considered the quick access to emergency functionalities during the ride to be useful, as it, e.g., supports a feeling of safety.

The map-based functionalities (e.g., navigation, context-based information) were positively mentioned by five experts. P1 and P4 especially considered the visualization of various ride options on the map to be helpful for understanding the differences between the offers. However, some experts (P2, P3) found the search results to be crowded and preferred fewer details. P3 would have liked to have all detailed information accessible on-demand, e.g., by expanding the journey overview.

While walking to the pick-up point, i.e., shortly before the departure time of the AV, P5 and P6 would want the UI to communicate how long the AV would wait for the passenger at the pick-up. P3 wanted the GUI to remember changes in filter settings for the next booking. In change of plans situations, P2 and P5 wanted the UI to be capable of making adjustments to already booked trips. However, in shared rides, this is conceptually restricted, as changes would also affect other passengers. When aborting a current trip, P2 would have expected that map-based functionalities are still accessible and that information on onward travel options is provided in order to proactively support the user. Overall, the GUI's visual design was described as appealing by five of the six experts. Further optimization potential was found regarding the design of single components and details (e.g., icon metaphors, button design).

Chatbot

The chatbot received excellent ratings for Stimulation ($M(SD) = 1.8(0.9)$), and Novelty ($M(SD) = 1.7(0.9)$) as well as above average ratings for Attractiveness ($M(SD) = 1.6(0.9)$), Perspicuity ($M(SD) = 1.7(0.8)$), and Efficiency ($M(SD) = 1.4(1.0)$), but a below average rating for Dependability ($M(SD) = 1.1(1.1)$) [222]. The latter indicates lack of predictability and feeling of control.

All five experts generally liked the conversational approach for interacting with the autonomous system. Three saw major advantages in the assistive and guiding nature of the chatbots' conversation flow and the high efficiency when users become experienced with the interaction. P2 especially appreciated the option to set all parameters at a time with a single message. However, this advantage was assessed to be less pronounced when users are not familiar with the chatbot. To counteract this issue, the experts suggested to optimize the on-boarding process.

All experts considered the chatbot's tone of voice to be appropriate to the bot's capabilities and role. P3 and P5 explicitly liked the neutral and friendly conversation style. However, two experts suggested to make the language even more human and to improve system feedback by using more relational messages.

P2, P3, and P4 recommended to combine chatbots with classic GUIs, i.e., to create a 'hybrid' UI with aspects of both. This idea was indirectly also emphasized by other experts since all participants would have wanted to always have access to common functions like, e.g., the map or tickets. P2, P4, and P5 would have wanted to directly manipulate set parameters in the messages sent by the bot as it is possible in classic GUIs. P1, P2, and P3 argued that the visual design of quick actions and buttons was not salient enough. Further optimization potential also concerned ambiguous or unclear wordings.

5.2.2 User Study

Descriptive and inferential statistics were calculated using JASP 0.12 [127], jamovi 1.2 [249], and R Studio 1.2 (R 4.0) [207]. Planned contrasts were performed to answer the hypotheses in terms of a potential underlying interaction effect. Repeated measurement ANOVAs were calculated to explore significant differences related to the used UI concept. Qualitative data from the interviews, session notes, and anecdotal evidence was digitally collected, structured, and analyzed.

Acceptance

Usefulness, Satisfying, Intention to Use, and Trust scales score high ratings in all conditions (Figure 5.4). While planned contrasts did not reveal the expected interaction effect in terms of Usefulness $(t(32) = -0.08, p = .398; d = -0.42)$ and Satisfaction $(t(32) = -0.17, p = .863; d = 0.08)$, a significant large interaction effect for Intention to Use $(t(32) = 2.99, p = .005; d = 1.45)$ was revealed. In other words, participants reported a higher intent to use the GUI in 'happy path' scenarios, but preferred the chatbot in scenarios with CoPs.

However, in general, the service was rated significantly more satisfying when using the GUI than when using the chatbot $(F(1, 32) = 5.25, p = .029; \eta^2{}_{\mathrm{G}} = 0.11)$. In terms of Usefulness and Intention to Use no such effect was found. The AMoD service also received high ratings for both UI concepts in terms of Trust without a meaningful difference between the scenario conditions $(F(1, 32) = 1.90, p = .178; \eta^2{}_{\mathrm{G}} = 0.053)$.

UX

Regarding the pragmatic quality, both concepts received high ratings in terms of Perspicuity, Efficiency, and Dependability (Figure 5.4). Planned contrasts did not reveal significant interaction effects – neither for Perspicuity $(t(32) = -0.04, p = .966; d = -0.02)$, Efficiency $(t(32) = -1.23, p = .227; d = -0.60)$, nor for Dependability $(t(32) = 0.26, p = .796; d = 0.13)$. While no meaningful differences in terms of the UI concept could have been found in the pragmatic quality scales, an ANOVA revealed significantly higher ratings for the GUI than for the chatbot and a medium effect in terms of Attractiveness $(M(SD) : GUI_{\mathrm{HappyPath}} = 2.0(0.6), Chatbot_{\mathrm{HappyPath}} = 1.4(0.9)$, $GUI_{\mathrm{CoP}} = 1.8(0.6)$, $Chatbot_{\mathrm{CoP}} = 1.3(0.8)$; $F(1, 32) = 5.56, p = .025$; $\eta^2{}_{\mathrm{G}} = 0.13)$. In other words, participants rated the overall impression of the service better when a GUI was used. Regarding the hedonic qualities Stimulation and Novelty no significant differences were found $(M(SD)$ for Stimulation: $GUI_{\mathrm{HappyPath}} = 1.0(0.9)$, $Chatbot_{\mathrm{HappyPath}} = 0.9(0.7)$, $GUI_{\mathrm{CoP}} = 1.5(0.7)$, $Chatbot_{\mathrm{CoP}} = 1.1(0.6)$; $M(SD)$ for Novelty: $GUI_{\mathrm{HappyPath}} = 1.1(1.0)$, $Chatbot_{\mathrm{HappyPath}} = 0.7(0.9)$, $GUI_{\mathrm{CoP}} = 1.1(0.8)$, $Chatbot_{\mathrm{CoP}} = 1.0(0.5))$.

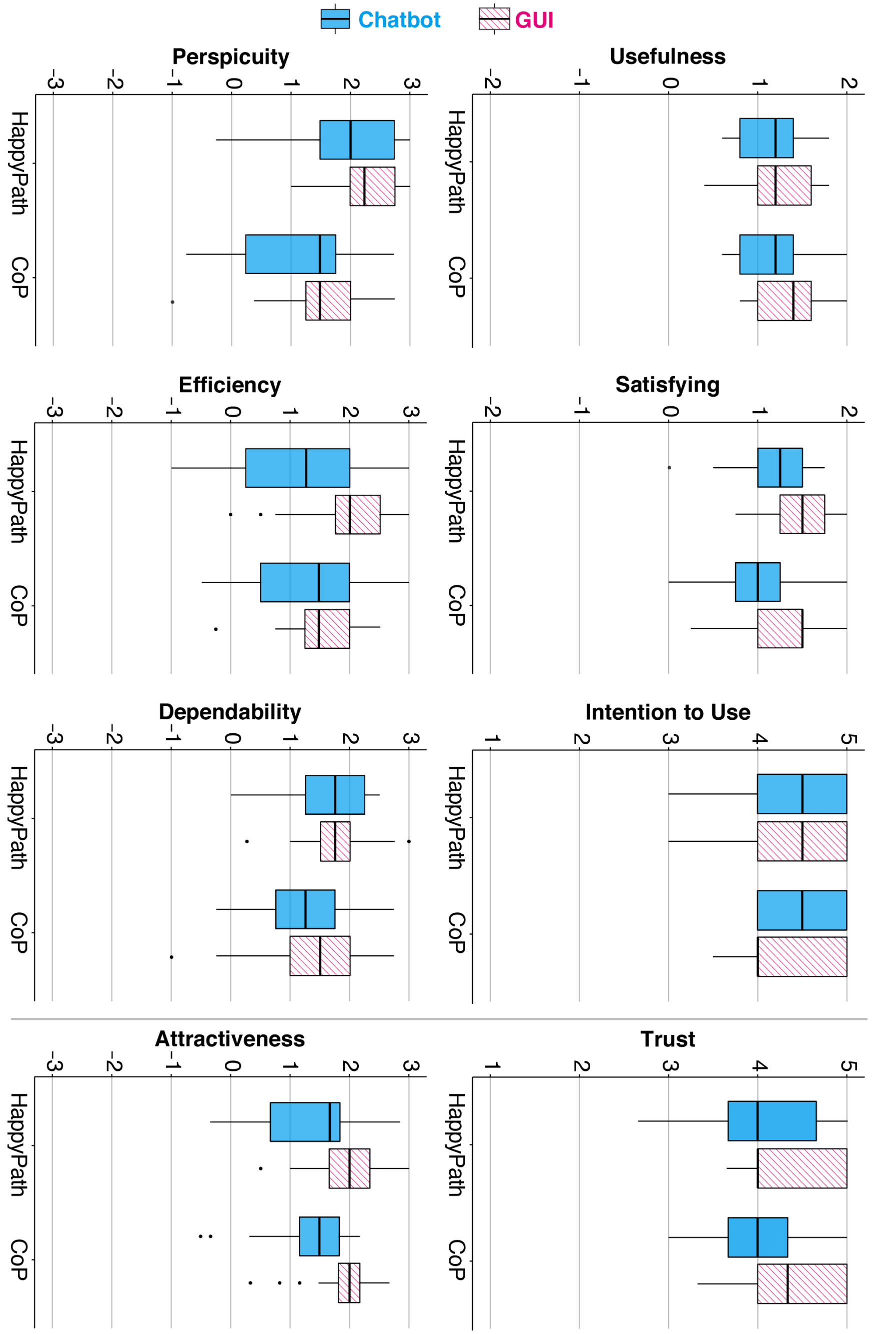


Figure 5.4 – Boxplots of acceptance and UX scales from the user study; $n_{\text{Chatbot}} = 17$; $n_{\text{GUI}} = 17$.

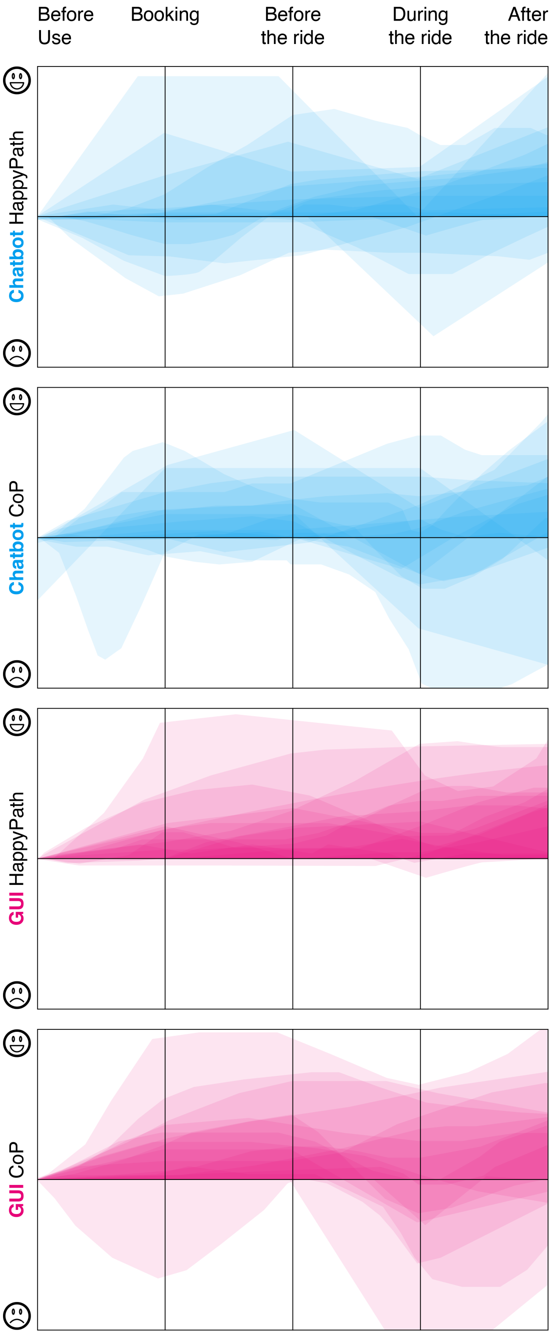


Figure 5.5 – Stacked emotion curves (opacity: 0.1) drawn by the participants after each journey, normalized at 'before use'. Top to bottom: chatbot 'happy path' $(n = 17)$, chatbot CoP $(n = 17)$, GUI 'happy path' $(n = 17)$, GUI CoP $(n = 17)$.

Emotion

Emotional curves indicate quite positive feelings throughout the journey and across conditions. For a comparative evaluation, the curves were digitized, normalized to 'before use', and stacked (opacity: 0.1) for each condition (Figure 5.5). The stacked curves reveal an overall positive emotion trajectory throughout the 'happy path' journey with the GUI. In the 'happy path' with the chatbot, they show larger fluctuations and some negative peaks during the ride and in the booking phase. During the booking phase and before the ride, almost all curves appear to follow a positive trend. For both the GUI and the chatbot, the trend seems to be slightly more positive in the CoP than in the 'happy path' conditions. I.e., changing the departure time seems to be a positive experience with both concepts. During the ride, participants indicate adverse effects in terms of their emotional constitution in all conditions except for the GUI's 'happy path'.

Presence, Wellbeing, and Simulator Sickness

IPQ scales scored medium to high ratings (Figure 5.6). While Involvement and Experienced Realism fell rather short with medium ratings, Spatial Presence and the General scale achieved above middle ratings (Figure 5.6). Four participants explicitly described the simulated ride as "realistic". P4, P17, and P28 commend the driving style of the simulated AV (e.g., "I feel relaxed", "I feel safe"). P2, P4, and P16 found it unrealistic that there was always a free parking slot available. Thirty of the 34 participants left the AV (i.e., the simulator) on their own when the simulation reached the target destination, while four remained seated. Similar to the IPQ's general scale, the single item *"I found the ride in the simulator realistic."* (Figure 5.6) achieved a high rating with a median of 4. Participants' did not seem to encounter adverse effects during the simulation, which is illustrated by a positive assessment of Wellbeing (median = 4) and low occurrence of Simulator Sickness (median = 1; see Figure 5.6).

Qualitative Findings: UI Design

Six GUI participants wanted to have a "direct" option to make adjustments to booked trips, e.g., change departure time without cancellation. P32 wanted the GUI to be more supportive ("take me by the hand") in CoP situations. Two participants preferred to have an option for speech-based in- and output. In general, some participants applied a 'trial and error" strategy on the GUI prototype, i.e., when they did not know what to do, they just tapped randomly.

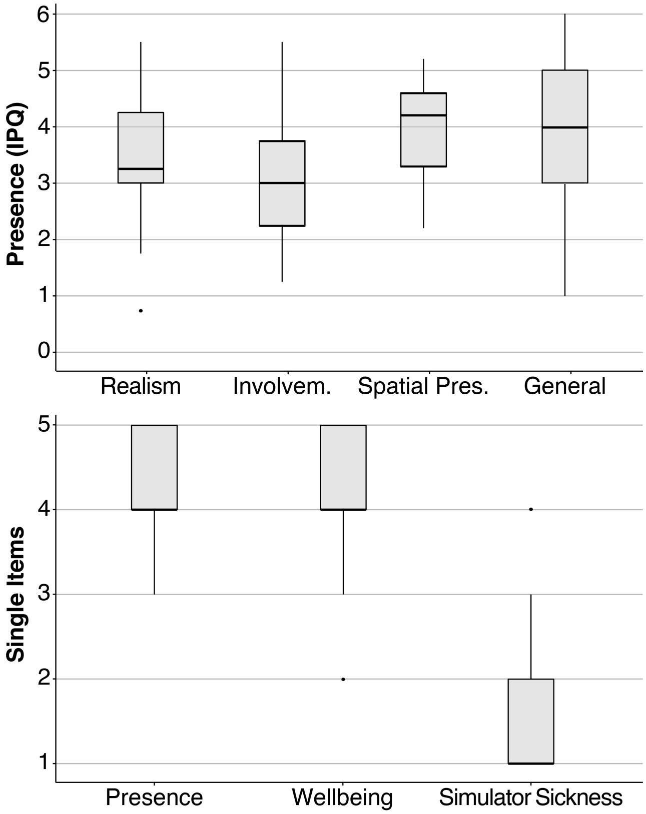


Figure 5.6 – Top: boxplots of the four IPQ scales Realism, Involvement, Spatial Presence and General (0 = min, 6 = max); bottom: boxplots of the assessed single items Presence, Wellbeing, and Simulator Sickness (1 = min, 5 = max); $N = 34$.

While all participants testing the GUI considered the type of the communication to be appropriate, three participants testing the chatbot either generally did not like bots (P4, P24) or found it inconvenient to articulate their needs using written text (P10). Several users (P12, P13, P23) feared that the chatbot would not understand them correctly if they would not use specific keywords. In contrast, P9 commended that the chatbot understands whatever they said. The qualitative assessment of the chatbot's tone of language confirmed the intended balance in terms of enthusiastic/matter-of-fact, formal/casual and funny/serious, as well as a rather respectful than irreverent way (see [172]).

Besides that, some participants would have wanted the chatbot to call them by their name (P10) and to make smalltalk (P12, P24). Five participants wanted the messages of the chatbot to be shorter – especially when they would use it more often. P11, P24, and P32 commended the chatbot's supporting nature, e.g., departure reminder. Five participants suggested to add direct touch interactions to the chatbot, e.g., to change values by clicking on chatbot messages. Further suggestions concerned speech output (P12), an always available "context view" (P7), and a app menu as used in classic GUIs (P4). Similar to the latter, P11 preferred a combination of both GUI and chatbot.

Qualitative Findings: AMoD Service Design and Simulation

Two participants (P25, P32) suggested to introduce gamification elements, e.g., a calculation of saved CO_2 using a shared AV compared to taking the own car. During the ride in the simulator 15 participants (7 using the GUI, 8 using the chatbot) stated that they did not have sufficient information. Seven of them (4 using GUI, 3 using chatbot) referred to a specific CoP situation where the in-vehicle display did not provide required information. This illustrates the crucial connection of various service touchpoints (e.g., mobile app, in-vehicle display) and the context.

Two participants (P13, P30) noted that they were bored during the simulated ride in the AV. P2 would have wanted to use his smartphones as they would typically do while riding in public transport or taxis. Several participants wondered how the AV would behave if, e.g., a passenger would not leave at the designated destination (P7, P17, P20), a passenger without a valid ticket entered the AV (P28), or an emergency occurred (P6, P28). P6 found it "embarrassing" if all other passengers in a shared AV would know who requested an extra stop.

5.3 Discussion

The conducted expert and user studies reveal the advantages and disadvantages of both chatbots and 'classic' GUIs – in general, and specifically for certain AMoD scenarios. Table 5.1 situates and contrasts these among the findings from previous work. In the following sections, we discuss how they relate to our research question regarding the UI concept's effect on acceptance and UX in 'happy path' and CoP scenarios. We also discuss the applied context-based prototyping approach [85], limitations, and future work.

Table 5.1 – Summary of general advantages (+) and disadvantages (–) of 'classic' GUIs and chatbots based on the findings of related work and our expert (E) and user (U) studies.

	'Classic' GUIs	Chatbots
+	High discoverability through visual clues [170], (U)	Intuitive input via natural language [164], (E), (U)
	Support users' mental model [170, 244], (E)	Expression of needs with own words [244, 245], (E), (U)
	Easy understanding of system scope [170, 244], (U)	Efficient for experienced users [244], (E)
	Suitable for displaying loads of information [37, 244]	Adaptability and personalization [91, 279], (E)
	Direct manipulation of elements (E), (S)	Ease of use [255], assistance and guidance (E), (U)
	Efficient shortcuts and established concepts [245], (E), (U)	Direct clarification of ambiguities [244]
	Potentially (more) attractive visual appearance (E)	Simplifiable information search process [255]
–	Users need to understand the design [244, 245], (E), (U)	Interpretation problems and misunderstandings [91], (U)
	Ambiguities of non-standard functions (E), (U)	Mentally demanding articulation of needs [170, 244], (U)
	Restrictive in terms of alternative usage ways (U)	Step-wise learning of system capabilities [244], (E)
	Potential information overload and distraction (E)	User requests can exceed system scope [245], (U)

5.3.1 GUI vs. Chatbot for AMoD Systems

Generally, both 'classic' GUIs and chatbots hold several advantages and disadvantages. As illustrated in Table 5.1, our results confirm related works' findings on GUIs and chatbots and complement the existing literature body with some new insights and a comparative case study on the two concepts for the AMoD domain. Across our studies, both the 'classic' GUI and the chatbot received high acceptance and UX ratings. Thus, the results confirm that both concepts can be regarded as valid and suitable options to consider when designing mobile applications for AMoD systems.

In the user study, the GUI was reported to be significantly more satisfying and attractive than the chatbot in general (both with medium-sized effects [46]). Descriptive plots of other acceptance and UX scales (Figure 5.4) support this tendency but do not reveal meaningful differences. The very positive assessment of the GUI's visual design in the expert and the user study might, however, be co-responsible for its higher Attractiveness ratings. While the chatbot prototype still achieved a relatively high assessment, it probably fell shorter in terms of visual appearance compared to the GUI because of restricted design possibilities. The general pros and cons of the concepts (Table 5.1) hold further possible explanations for the effects.

Planned contrasts revealed a significantly higher Intention to Use the GUI in 'happy path' scenarios but the chatbot in CoP scenarios. This is partly supported by the GUI's emotion curves (Figure 5.5). They revealed positive experiences throughout the 'happy path' scenarios with the GUI but relatively large derivations in CoP scenarios. However, the chatbot's emotion curves (Figure 5.5) and other acceptance scales (Usefulness, Satisfying, Trust; Figure 5.4) do not back this observation. Consequently, we have to reject the acceptance-related hypotheses H1.1 and H1.2. Similarly, we have to reject H2.1 and H2.2 as none of the UEQ's pragmatic quality scales returned the expected interaction pattern.

In both chatbot studies, participants preferred to directly manipulate parameters (e.g., date, time, destination) displayed by the bot. While this can be regarded as a clear benefit of GUIs in general (Table 5.1), it should be considered whether it might also be a usable option to implement direct manipulability in chatbots – e.g., form elements, date pickers, app menus. Alternatively, since several non-standard functions and information in the GUI have not been 'instantly' understood in both the expert study and the simulator user study, we recommend supplementing the 'classic' GUI with conversational elements. The assistive nature of chatbots (and CUIs in general), their linear conversation flow, and the on-demand provision of adequate explanations seem to be an excellent addendum to overcome the associated challenges of new and non-standard functions for future AMoD systems.

5.3.2 Context-based Prototyping with Immersive Video

As AMoD is still a theoretical matter [193], we considered simulating holistic user journeys as a crucial component for successful evaluation and, thus, for developing adequate human-AV interactions that can counteract acceptance hurdles.

The video-based approach adapted from [85] generally provided a straightforward and controllable environment that made it possible to immerse participants in the context in order to let them experience a complete AMoD journey. Assessment of the IPQ scales Realism, Involvement, Spatial Presence, and its General scale supports this with medium to high scores, along with qualitative findings. Although our TV-based setup differed from the projection-based setup of [85], the results are quite comparable. This indicates a successful replication of the approach. The very positive assessment of the single items on presence and wellbeing as well as the low occurrence of simulator sickness symptoms further supports the general suitability of the method.

Since people seemed to become bored doing nothing during the (simulated) ride – which might have affected participant's assessment, e.g., in terms of emotion curves – it might be beneficial for further empirical studies to introduce secondary tasks (e.g., attention tests, monitoring of changes in the UI, or simply reading a book).

5.3.3 Limitations

Some limitations of this work should be noted. In general, this work focuses on mobile applications for AMoD. Consequently, the generalisability of the reported studies and the comparative evaluation of GUIs and chatbots is limited.

Created prototypes were optimized for following a pre-defined scenario. However, both prototypes featured alternative interaction paths to some extend. E.g., in the CoP scenario 'change of destination', the chatbot offered a step-by-step approach similar to the GUI but was also able to provide a context-based shortcut if the participant directly requested to change the destination via text. While such alternative usage ways incorporated the concepts' benefits into the study, they might also affect the study's reliability.

The GUI was created using Sketch's prototyping features which highlighted interactive elements on tap. As a result, participants applying a 'trial and error' procedure received hints from the prototype on how to proceed. The chatbot prototype made use of the language understanding capabilities provided by Botmock. Though we added an extensive collection of likely intents and formulations, not all possible formulations were anticipated. Thus, some participants encountered 'dead ends' during the study and had to try different formulations. In some cases, this resulted in a trial and error behavior as well. Consequently, both prototypes' levels of fidelity might have impacted the results.

Participants in all three studies reported a rather high affinity for technology interaction. While this is considered to be a common phenomenon in HCI studies [94], it might impair external validity. Furthermore, having a look at the very positive evaluation of both prototypes and scale constitutions, ceiling effects cannot be excluded.

The user study was conducted in the late summer of 2020, i.e., during the COVID-19 pandemic. Following regulations and recommendations of local and national authorities, we applied several precautions, such as distancing and hygiene measures. While we consider the effects on the study's actual conduct to be minor, the situation might have affected the sample composition, as, e.g., only people with low fear of the situation might have signed up for the study.

5.3.4 Future Work

Similar to the recommendation of [92] to integrate relational and productivity-oriented interaction for conversational design, future work should consider transferring this approach in terms of combining GUI-based with CUI-based concepts to create flexible and accessible interactions. Such could be text-based but, considering accessibility requirements, also speech- or gesture-based. Like 'hybrid' UIs in today's cars, the approach could combine the benefits of both worlds.

Future work should investigate appropriate use cases and create design guidelines for such 'hybrid' AMoD UIs. In doing so, different in- and output modalities should be considered – also making use of further technological advances like text-to-speech conversion and vice versa. In terms of the system's accessibility, this could be used to integrate specific demographic groups' needs – e.g., blind people or older adults might feel more confident talking to an AMoD system instead of chatting or tapping.

Furthermore, as some participants suggested, integrating gamification elements (e.g., calculating saved CO_2) might hold attractive potentials for future designs to increase the system's acceptance and hedonic quality.

5.4 Conclusion

Both GUIs and chatbots come with their specific pros and cons. Based on our findings, we derive three design recommendations for human-AMoD interaction. (1) Use ‘classic’ GUIs as a basis for mobile interaction. In direct comparison, the GUI was reported to be generally more attractive and satisfying. GUIs also seem to be superior in standard use cases with ‘happy path’ scenarios. However, the conversation with a chatbot-based agent can increase users’ intention to use the system and support users in demanding scenarios. Therefore, we recommend to (2) integrate (text-based) conversational elements in GUI-based mobile applications where appropriate to enhance user control and facilitate error recovery. We further recommend to (3) use context-based prototyping [118, 85] from early development phases on to consider environmental factors and the interdependence of various touchpoints (e.g., mobile apps and in-vehicle displays). Video-based simulation [85] provides a suitable and straightforward basis to consider all stages of the AMoD user journey holistically. Future work should investigate applying these recommendations to other usage scenarios and the creation of design guidelines for accessible AMoD systems.

5.5 Chapter Summary

As a foundation for deriving empirically grounded design guidelines for human-AMoD interaction, this chapter investigated two approaches for mobile interaction with AMoD systems: chatbots and ‘classic’ graphical UIs (GUIs). We evaluated prototypes of both in expert studies ($n_{\text{GUI}} = 6$; $n_{\text{Chatbot}} = 5$) and a between-subjects simulator user study ($N = 34$). For the latter, we adapted the immersive video-based simulation approach presented in Chapter 3. In contrast to the initial setup, we replaced the video projectors with TVs. Furthermore, we enhanced participants’ spatial immersion with wooden pallets to create an elevated platform which was embraced by the screen configuration and an additional whiteboard that closed the space from the back. In the user study, we used the setup to let participants experience a complete AMoD user journey in a controllable lab-based setup encompassing interactions before, during, and after the ride. The adapted setup received similar assessments regarding participants’ presence perception, which supports the suitability of the cost-effective method and hints at a successful replication of the approach proposed in Chapter 3.

The compared concepts for mobile interaction with AMoD systems received both good ratings in terms of acceptance and UX. The 'classic' GUI concept results in significantly higher attractiveness and user satisfaction ratings than the chatbot. A significant interaction effect revealed a higher intention to use the chatbot in scenarios involving a change of plans, while the GUI was preferred in 'happy path' scenarios. These quantitative findings were supported by our qualitative interview data and the emotion curves drawn by the participants after each ride. Based on our findings and a comprehensive review of related work, we provided a summary of the advantages and disadvantages of both concepts and accordingly derived design recommendations. We concluded with recommendations to consider GUIs as a basis for mobile interaction with AMoD systems, but to supplement these with (text-based) conversational components to enhance user assistance, notably to support unfamiliar functions and scenarios such as occurring change of plans.

Regarding our general research questions, this work contributes to **RQ1** by investigating and comparing two concepts for mobile interaction with AMoD systems, namely 'classic' GUIs and chatbots. Through the empirical investigation of Q5.1, we derived a comprehensive overview of the advantages and disadvantages of the two concepts and design recommendations for AMoD HMIs. Both researchers and practitioners can build on the findings to design suitable HMIs that counteract acceptance challenges. Furthermore, contributing to **RQ2**, we demonstrated how the video-based simulator approach proposed in Chapter 3 can smoothly adapt to different hardware setups. Furthermore, we showed how such a lab-based setup could be used to simulating a complete AMoD user journey in a controllable environment encompassing activities before, during, and after the ride with a (shared) AV.

6

Passengers’ Information Needs in Shared Day and Night Rides

In shared AMoD systems, passengers will share rides with strangers during both day and night times. Since there is no human authority (e.g., a bus driver) present anymore in AVs, users’ acceptance is likely to be influenced by the presence of co-passengers [212, 227]. Information about fellow travelers prior to and during the ride seems to have the potential to affect user acceptance positively [139]. The question arises whether the findings of König et al. [139] are further influenced by gender and time of the day as, e.g., [227] found that especially women are feeling anxious to share rides at night times.

We agree with Schuß et al. that "research is needed on how to enhance women’s security while not leaving other groups of people out" [228, p. 12]. Research in

---

THE CONTENTS OF THIS CHAPTER WERE PUBLISHED PREVIOUSLY AS PART OF:

[87] **Lukas A. Flohr***, Martina Schuß*, Dieter P. Wallach, Antonio Krüger, and Andreas Riener. 2023. Designing for Passengers’ Information Needs on Fellow Travelers: A Comparison of Day and Night Rides in Shared Automated Vehicles. *arXiv* 2308.02616 (aug 2023), 22 pages. `https://doi.org/10.48550/arXiv.2308.02616`.
**Note: both authors contributed equally to this paper.*

public transportation shows that feelings of anxiety and unease when traveling with strangers greatly influence perceived security [57]. Building upon related work, we hypothesize that knowledge about co-passengers might positively affect perceived security, acceptance, trust, and overall UX. Consequently, this chapter investigates the following question and, thereby, contributes to our general research question **RQ1**.

Q6.1 How are shared autonomous mobility-on-demand (SAMoD) passengers' perceptions of security and corresponding UX, trust, acceptance, and emotions influenced by the time of day of a ride and their knowledge about co-passengers?

Most related work on SAMoD used (online) surveys to examine user acceptance of (shared) AVs. As such, participants lack context, i.e., experience using SAMoD systems. Therefore, we investigate our research question in a user study using a controlled simulated environment, enabling this very experience. On this basis, our empirically grounded findings confirm but also challenge previous works on shared AVs and shed new light on the effects of time of day and the interrelation of (co-)passengers.

6.1 Material and Method

To investigate our research question, we created a UI prototype representing a SAMoD in-vehicle passenger information display. Variants of the UI were evaluated in an exploratory within-subjects user study with a diverse sample of participants ($N = 24$; gender-balanced, wide range of ages) using a video-based automated vehicle simulator [85]. With our study, we aim to have a closer look into the information needs of passengers and counteract the limitations of an online survey by simulating rides in an SAV during different times of the day. At the same time, we do not necessarily say this would be the best solution. Quite the contrary, we acknowledge that 1) serious privacy side effects could arise, and 2) stereotypes could be further manifested. Still, we wanted to let participants experience receiving this information about their fellow travelers and discuss with them how it would influence their perceived security. Our motivation was to evaluate whether such a controversial concept would convey security after all and, if so, under what circumstances people would need and want to use it.

We identified leisure trips as a typical case for the use of SAMoD during both day and night times (more details in section 6.1.3). In each simulated ride, an in-vehicle

Figure 6.1 – Study participants experienced two day and two night rides in the immersive video-based automated vehicle simulator.

UI prototype (section 6.1.2) provided participants with information about the ride and fellow passengers. In the respective UIs, we varied the type and amount of information participants received when co-passengers boarded and left the vehicle. Information was either provided *with* personal data on co-passengers (name, age, target destination, profile picture), or *without*. We used a within-subjects design, and each participant experienced four rides: two night and two day trips, one with and one without personal information about co-passengers. The order in which each participant experienced the variants was randomized and counter-balanced.

Since we wanted to investigate shared rides, we identified two options to include the "sharing" aspect in the study: 1) simulating passengers boarding/leaving the vehicle with sounds and visual information displayed on the UI, and 2) using real persons ('actors') that complement the setup. Regarding the latter, Flohr et al. [85] investigated the effect of supplementing SAV simulator studies with actors mimicking co-passengers. While they found some support for the approach, it does not seem to increase participants' immersion in the simulation. Instead, it seems to increase the occurrence of motion sickness symptoms in simulator studies [85]. Therefore, considering the potential adverse effect on participants' wellbeing during the simulator study and the problematic pandemic situation at the time of the study conduct, we decided to simulate co-passengers getting on and off the SAV only virtually. While this supports, on the one hand, our intended focus on the information display, this can, on the other hand, also be considered a limitation of the study, which we further discuss in section 6.3.4.

The study was conducted in accordance with the ethical guidelines stated in the Declaration of Helsinki [275]. Participants took part voluntarily, were obliged to provide their written informed consent, and had the opportunity to abort the study at any time without stating reasons.

6.1.1 Setup

Since contextual factors play a crucial role in passengers' travel experiences and information needs, we intended to establish a realistic but still controllable test environment for the user study. Therefore, we adapted the immersive video-based simulation setup used by Flohr et al. [85, 86] (Chapters 3 and 5) and combined it with a tent-based vehicle mock-up (e.g., used by Schuß et al. [228]) to provide even more realism. The resulting setup (Figure 6.1, 6.3) consisted of three LCD screens that played back videos representing a passengers' view out of the front, left, and right windows of a shared AV.

Similar to [85], we used audio and video footage of day and night rides through an urban environment to create simulations for two night and two day rides. The footage was captured using three action cameras mounted in the center of a BMW i3's windshield, as well as on the front side windows. In addition, we enhanced audio footage with additional sounds (e.g., opening and closing noises of sliding doors). Along with a 2x2 seating group, the footage was played back on three NEC Full HD 55.1-inch TV screens situated in a tent-based vehicle mock-up. The tent separated the simulation from the surrounding lab environment to support participants' immersion by entering a closed space when boarding the simulated SAV (Figure 6.3). The UI prototype of the passenger information display was displayed visually on an additional 24.1-inch screen (Figure 6.1). Audio sounds and voice prompts were provided by a Logitech 2.1 sound system.

6.1.2 Design Process and Prototypes

The tested UI prototypes were designed iteratively following findings from related user studies and a comprehensive literature review. We used video-based prototyping to create high-fidelity visual and auditory UI representations that matched the video-based simulation. The visual information display featured a split-view of 1) a schedule showing upcoming stops, estimated arrival times, and information on co-passengers getting on/off the vehicle, and 2) a map illustrating the current location of the AV and the planned route (Figure 6.2), which follows proposals of previous work (e.g., [213, 85]). We created two general prototype variants to investigate the research question (Figure 6.2). While the first variant ("without") does not show personal information about co-passengers, the second variant ("with") features such information by displaying

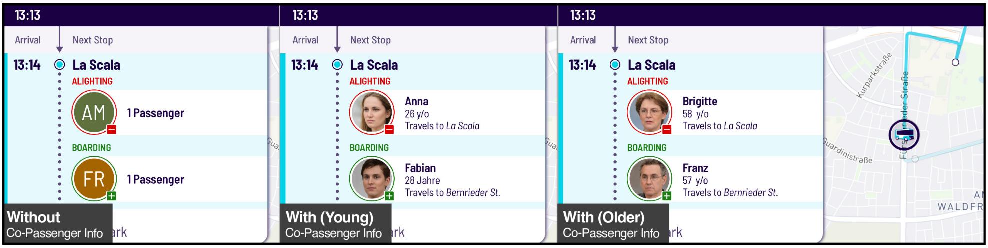


Figure 6.2 – Apart from time of day ("day" and "night"), study conditions varied in the amount of provided information on co-passengers: 1) without information, 2) with information. In the two rides with information, co-passengers' age varied between "young" and "older".

name, age, target destination, and profile picture of co-passengers. For each test ride, participants experienced either a prototype with or without information on co-passengers, i.e., the variant stayed consistent within the rides.

Previous research suggests that combining these data reduces overall compensation demands for sharing a ride with a stranger [139]. We did not include a rating of fellow passengers, as rating systems hold discriminating characteristics [21, 228]. We used AI-generated pictures with neutral facial expressions [97] as photos of the entering fellow passengers. We included fellow passengers' age as we hypothesized that this information might influence participants' perceptions. Thereby, we defined two age groups: young (between 20 and 30) and older (between 50 and 60). Age of fellow passengers was balanced so that each participant experienced one ride with a younger man/woman and an older man/woman as we expected that age could have an effect on passengers' perceived security.

The provided contextual information (map, street names, etc.) matched the real-world environment where the simulation footage was recorded and animated (using Adobe After Effects CC 2021) according to the simulated vehicle's movements (e.g., the position of the AV in the map). For permutation purposes, we created eight video prototypes of the UI to have one variant without and one with information on co-passengers for all four simulated rides. Signal sounds and voice prompts complemented the visual UI (e.g., without: *"Next stop: [stop name]. One passenger gets on. One passenger gets off."*; with: *"Next stop: [stop name]. [Name of passenger] gets on. [Name of passenger] gets off."*). Voice prompts were created using text-to-speech conversion by Microsoft Azure.

6.1.3 Scenarios

We intentionally included participants covering a wide age range in the study, with young people not working yet, and older adults who do not work anymore. To provide for a broad spectrum of participants’ real lives, we chose leisure trips as scenarios for the four rides in the study. Since people are reported to be more likely to reject sharing rides with unknown fellow passengers for leisure trips compared to commute trips [151], we wanted to explore whether information about other passengers would mitigate this observation. All participants engaged in four trips: two during the day and two at night. We used storytelling to create authentic scenarios for each trip to enhance immersion. The day trips went from a bakery to a park to meet friends and back. The night trips started nearby the passenger’s home and had a restaurant as a destination where some friends were supposed to meet and were also round trips. To become even better acquainted with the scenario, participants received a paper ticket before each ride with their name, destination, departure and arrival time. After reading the scenario to them and handing over the ticket, our participants entered the shuttle bus, chose one of the seats in the front row, and one of the investigators started the video simulation. During each trip, one man and one woman as a co-passenger entered the vehicle virtually (i.e., this was only stated by the information displayed in the UI prototype). We did not randomize the order, i.e., it was always the woman entering first to avoid losing statistical power due to too many conditions. However, participants always rode with only one person at a time since we hypothesized that it would affect participants’ perceived security whether they would be sharing rides with a single man/woman or multiple persons simultaneously. The first (virtual) co-passenger entered at the first stop and got off at the second stop, where the second co-passenger entered the vehicle. At the third stop, participants’ reached their target destination.

6.1.4 Procedure and Measurements

We used a mixed-method approach [52] and triangulated quantitative data collected during and between rides with observations and qualitative interview data. Each study session can be divided into three parts: briefing and pre-questionnaire, test rides and measures, and post-session interview.

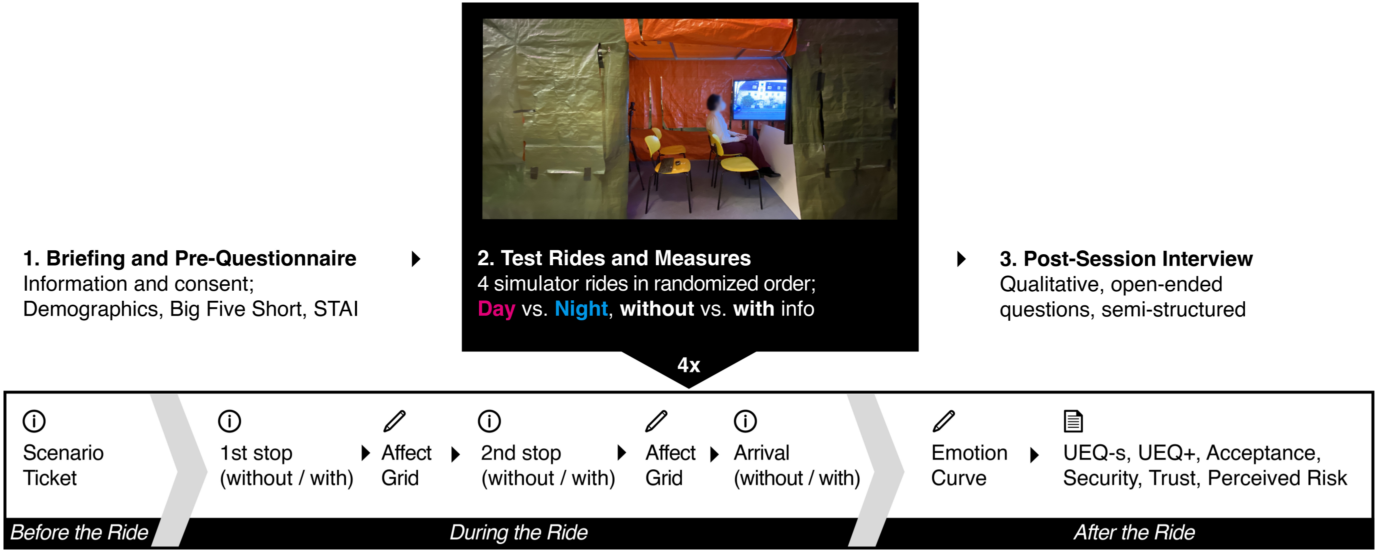


Figure 6.3 – Study procedure (top) and sequence of the four simulated SAMoD rides (bottom).

Briefing and Pre-Questionnaire

After receiving a briefing comprising general information about the study goal and the procedure, participants signed a declaration of consent. Then, they filled out a demographic pre-questionnaire. We also included the short version of the Big Five inventory [199] to get insights into a participant's personality. Prior research showed that psychological factors and attitudes most likely influence people's adoption of AVs [276, 106]. As the level of a person's anxiety influences the perceived security [147], we also included the state-trait anxiety inventory (STAI) [237] in our pre-questionnaire. Since current research is not conclusive on whether having experienced any sort of crime has an influence on perceived security [57], we left this aspect out.

Test Rides and Measures

During each of the four rides, participants filled out Russell's Affect Grid [209] in an adapted emoji-based version inspired by [252] using pen and paper. The Affect Grid is one of the most widespread models for emotion measurement and consists of two dimensions to measure: pleasure (displeasure – pleasure) and arousal (low energy – high energy). Each time information about an upcoming stop and entering or leaving passenger was displayed during the ride, participants were instructed to set a cross to express their current emotional state in the grid. After each ride, participants got off the simulated automated vehicle and summarized their subjective emotional constitution throughout the journey by drawing an emotion curve on a template also used by [135, 86]. Subsequently, the experimenter accompanied them to a workplace where they filled out a digital questionnaire. Starting with the short version of the User Experience Questionnaire (UEQ-s; 8 bipolar items; 7-point scale; [222]) as well as the Usefulness [223] and Attractiveness [150, 220] dimensions of the UEQ+ (4 bipolar items for each dimension; 7-point scale; [223]) participants assessed their experiences of the ride and respective HMI concept. Since we expected the type and amount of provided information to have an effect on passengers' trust, participants also assessed the Trust in Automation scale of Körber (2 items; five-point Likert-type scale; [141]). Furthermore, we investigated users' acceptance with the Intention to Use (2 items; 5-point Likert-type scale), and Perceived Usefulness (3 items; five-point Likert-type scale) dimensions of Chen's adaption of the technology acceptance model [41]. Subsequently, we included Dekker's Security Concerns scale (1 item; 5-point Likert-type scale; [62]) and the Perceived Risks scale (1 item; 5-point Likert-type scale; [201]) as risk also has an influence on the perceived security [147]. After the last ride, each participant additionally filled out the Igroup Presence Questionnaire (IPQ, 14 items; 7-point Likert-type scale; [224, 225]) to assess the quality and immersion of the simulated environment.

Post-Session Interview

Finally, we conducted a semi-structured post-session interview with each participant. We asked open-ended questions about the rides in general and the co-passenger information that was provided by the UI. Participants were asked which version of the UI they liked best and why. Participants were also prompted about potential feelings regarding security in the respective conditions, and we inquired whether some information was missing from their point of view. With the consent of participants, audio captures of all post-session interviews were recorded for an in-depth post hoc analysis.

6.1.5 Participants

In total, 24 participants (12 women, 12 men, 0 diverse, 0 n/a; from 18 to 81 years, $M = 40.5$, $SD = 21.3$) took part in the study. All participants were recruited through university mailing lists and word of mouth and attended the study voluntarily. For participation, all of them received financial compensation (approx. 25 US dollars). Their national background was German, Indian, Turkish, Russian, and Iranian. We used the STAI inventory to measure participants' interindividual tendency to evaluate situations as threatening or to react with increased feelings of anxiety. According to the reference values of the trait anxiety scale (items 21-40; [237]) our participants are at the expected medium level of responding with anxiety. The women in our study had a mean value of $M = 36.5$ ($SD = 7.0$; expected value according to references = 37.0) and the men a mean of $M = 35.8$ ($SD = 4.4$; expected value according to references = 34.5). Participants fall into the average age group between 36 to 65 years and have a high educational level. They correspond approximately to the reference values of the Big5-short [200] for extraversion ($M(SD) = 3.25(1.29)$; reference: $M(SD) = 3.62(.91)$), agreeableness ($M(SD) = 3.56(0.9)$; reference: $M(SD) = 3.43(.79)$), conscientiousness ($M(SD) = 3.93(1.03)$, $M(SD) = 3.47(.95)$; reference: $M(SD) = 4.2(0.77)$), neuroticism ($M(SD) = 2.45(0.94)$, $M(SD) = 2.48(0.9)$) openness to experience ($M(SD) = 3.45(1.21)$; reference: $M(SD) = 3.70(0.89)$). We therefore assume that the obtained results are not falsified through a non-representative sample (e.g., a sample with exceptional high scores in neuroticism could have an impact on the perceived security).

6.2 Results

For the quantitative results, descriptive and inferential statistics were calculated using JASP 0.16 and jamovi 2.2.5. The audio-recorded post-session interviews were transcribed verbatim and analyzed applying qualitative content analysis [29, 163] with MAXQDA. Session notes and anecdotal evidence during the study complemented the data collection.

6.2.1 Dependent Variables

In this section, we report on descriptive and inferential statistics for a comparison of the study conditions in terms of our dependent variables, as well as for an assessment of the simulated setup by having a look at participants' presence perception. We computed repeated measures analysis of variances (RM-ANOVA) to explore differences in the study conditions with the RM factors ′time of day′ (day, night) and ′information on fellow passengers′ (without, with) as well as the between subjects factor ′gender′ (women, men). One woman (P21) only completed three of the four rides due to occurring simulator sickness symptoms. The missing data of P21 was imputed with maximum likelihood estimates (e.g., [4]) for the respective scales. When a RM-ANOVA returned significant ($\alpha = .05$) for a certain scale, post-hoc tests in the form of Holm-adjusted pairwise comparisons for all conditions were calculated. Effect sizes were interpreted according to Cohen [46].

UX

With reference to the UEQ-s benchmarks [221, 222], the tested SAMoD system received excellent ratings for both pragmatic and hedonic UX quality throughout study conditions (Figure 6.4). While we did not find meaningful differences in terms of pragmatic quality , hedonic quality , and usefulness, a RM-ANOVA revealed significant differences for the UEQ's attractiveness scale with regard to time of day ($F(1, 22) = 6.820, p = .016, \eta^2{}_{\mathrm{G}} = 0.026$) and an interaction effect of passenger information and gender ($F(1, 22) = 5.059, p = .035, \eta^2{}_{\mathrm{G}} = 0.021$). Post-hoc tests show that participants' overall impression was significantly more positive ($t = 2.612, p_{\mathrm{holm}} = .016$) during daytime than during nighttime (Figure 6.4), with a mean difference of $M(SE) = 0.3(0.1)$ and a medium effect of $Cohen's\ d = 0.533$.

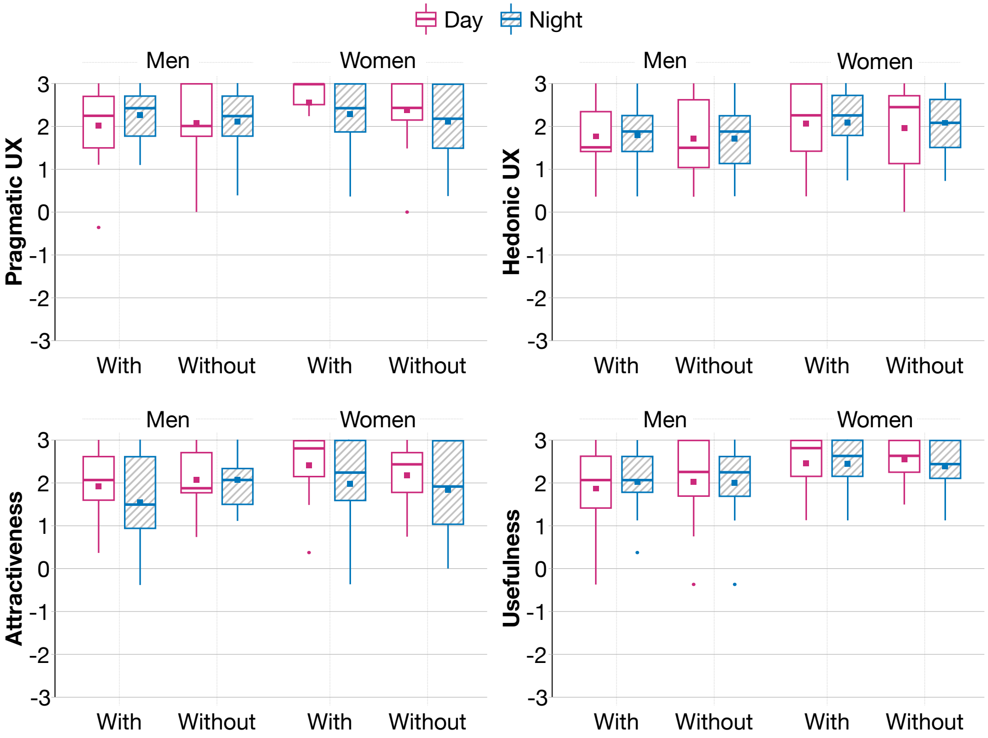


Figure 6.4 – Boxplots of UEQ-s scales (pragmatic UX and hedonic UX), usefulness, and attractiveness (-3 = low; 3 = high) for the four study conditions and the between-subjects factor gender.

Despite the significant results of the RM-ANOVA, an interaction effect of information and gender was not confirmed by subsequent pairwise comparisons.

Acceptance

A between-subjects effect of gender returned significant in the RM-ANOVA for both used scales of Chen's TAM [41]: Perceived Usefulness ($F(1, 22) = 7.586, p = .012, {\eta^2}_{\mathrm{G}} = 0.194$) and Intention to Use ($F(1, 22) = 6.490, p = .018, {\eta^2}_{\mathrm{G}} = 0.159$). Post hoc comparisons confirm that women perceive the tested SAMoD system to be more useful than men do (Figure 6.5; $t = 2.754, p_{\mathrm{holm}} = .012$) with a mean difference of $M(SE) = 0.5(0.2)$ and a medium-sized effect of $Cohen's\ d = 0.562$. Similarly, women show a higher Intention to Use the SAMoD system compared to men (Figure 6.5) with a mean difference of $M(SE) = 0.5(0.2)$ and a medium-sized effect ($t = 2.547, p_{\mathrm{holm}} = .018$,

$Cohen's\ d = 0.520$). Apart from the between-subjects effect and the generally medium-high to high acceptance ratings of the SAMoD system, no meaningful within-subjects effects of time of day and passenger information on Perceived Usefulness and Intention to Use were revealed.

Security, Trust, and Perceived Risk

Participants' trust in the automated system was medium-high among all conditions (Figure 6.5). With regards to the medium-rated security concerns (Figure 6.5), participants seem to have some, but no severe concerns on their security during their ride. No meaningful difference induced by time of day or passenger information was detected. A significant difference was found in terms of perceived risks ($F(1, 22) = 7.321, p = .013, \eta^2{}_{\mathrm{G}} = 0.013$). AMoD rides without information were perceived as significantly more risky than rides with information about fellow passengers (Figure 6.5) with a mean difference of $M(SE) = 0.2(0.1)$ and a medium-sized effect ($t = 2.706$, $p_{\mathrm{holm}} = .013$, $Cohen's\ d = 0.552$).

Emotion

Judging from visual inspection of the affect grids and emotion curves (Figure 6.6), participants found rides during daytime and without information to be most pleasant. Rides without information seem to receive more positive assessments whereas the UI variants with information show higher dispersion in the affect grids. Generally, rides during daytime seem to be perceived more pleasant than night rides. In accordance with that, the statistical analysis of the quantified ($min = 1, max = 10$) uni-dimensional subscales of the affect grid (pleasure, arousal) revealed no meaningful effect in terms of arousal but significant differences in the pleasure ratings with regards to the time of day ($F(1, 43) = 12.386, p = .001, \eta^2{}_{\mathrm{G}} = 0.032$). Rides during daytime ($M(SD) = 7.6(2.2)$) received higher pleasure ratings than rides during nighttime ($M(SD) = 6.8(2.2)$) with a mean difference of $M(SE) = 0.8(0.2)$ and a medium-sized effect ($t = 3.519, p_{\mathrm{holm}} = .001$, $Cohen's\ d = 0.525$).

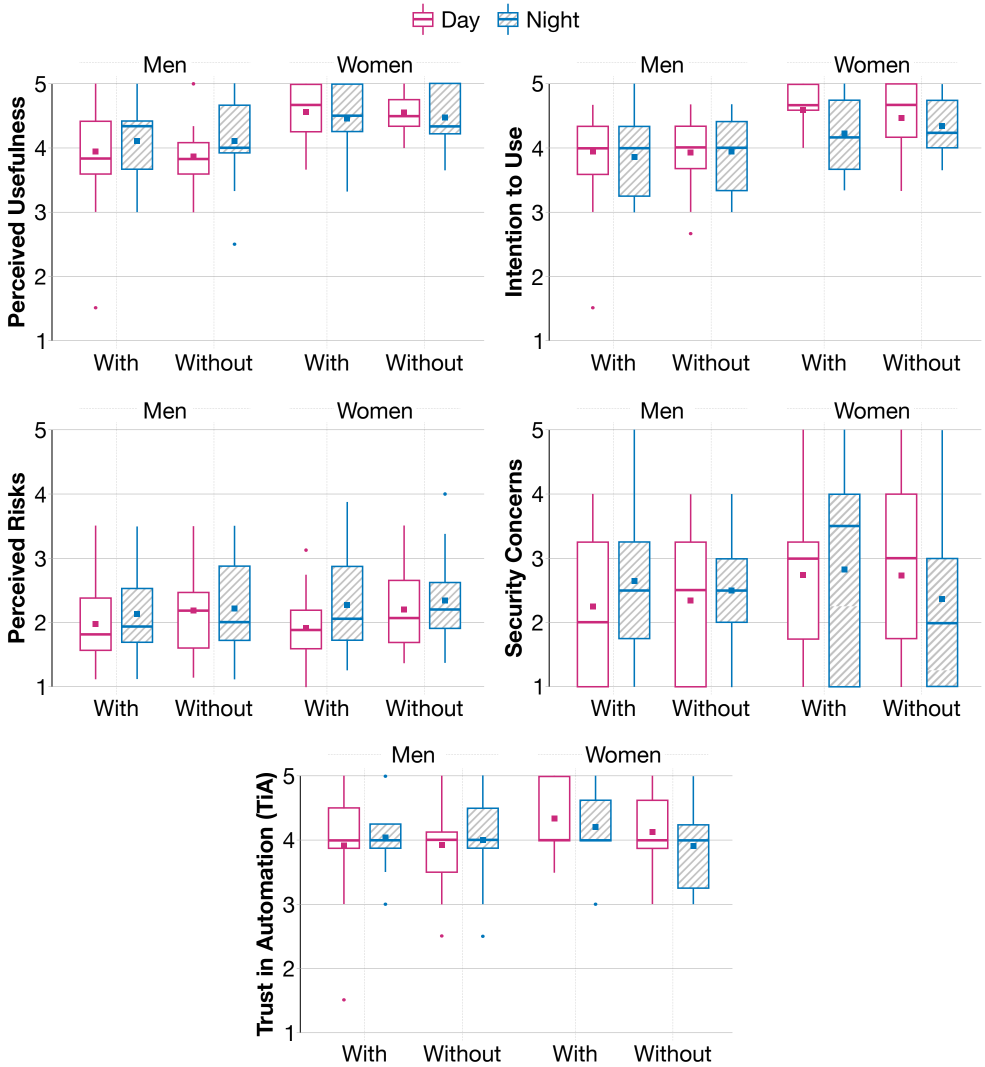


Figure 6.5 – Boxplots of acceptance scales (perceived usefulness, intention to use), trust in automation, security concerns, and perceived risk (1 = low; 5 = high) for the four study conditions and the between-subjects factor gender.

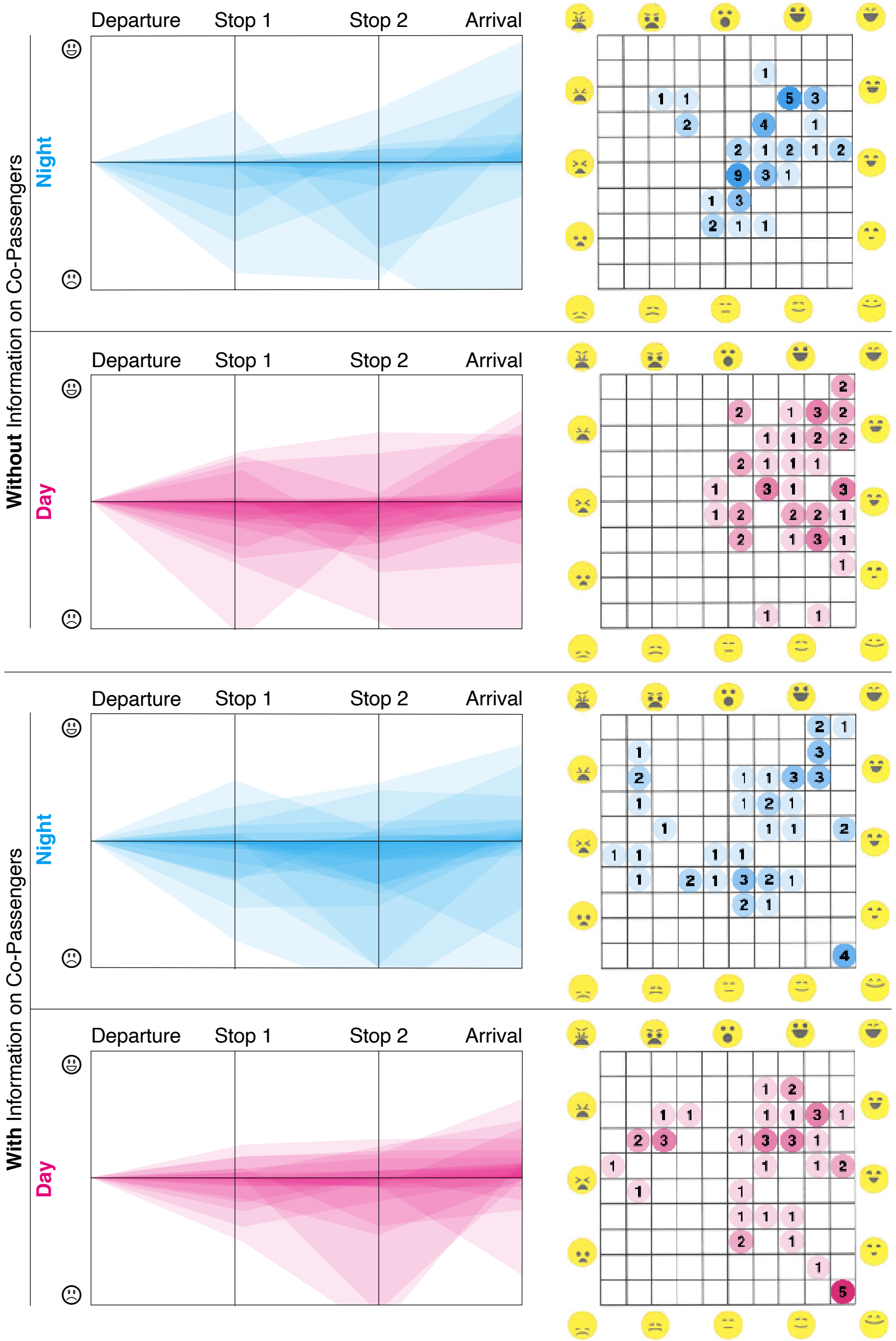


Figure 6.6 – Stacked emotion curves (left; opacity: 0.1; normalized at 'departure') and affect grids (right) for the four study conditions.

6.2.2 Qualitative Content Analysis

For the qualitative content analysis [29, 163], interview transcripts were initially explored line-by-line. In a second step, we highlighted text passages, searched for keywords, and added notes. Subsequently, the transcripts were scrutinized again and codes were derived from the text by applying inductive coding to refine themes and codes in an iterative process until the final expressions were identified. Below, we present our main findings (e.g., statements expressed during the post-session interviews) with their number of mentions and the number of women and men in our study mentioning them. First, we present the perceptions of the rides in general. Then, we cluster them according to three main topics: information preferences, day vs. night, and the type of information that participants were requesting.

Presence Perception and Experience of the Rides

In general, participants described the four rides as positive and considered the ride in the simulator as short, entertaining, and pleasant. Moreover, participants emphasized how realistic the four trips felt to them: *"Yes, it was quite real and I didn't feel I am in the simulation room and it was so real. It was quite good, yeah."* (P15), which is also reflected in the medium to high ratings for the four IPQ scales (Realism ($M(SD) = 4.0(1.1)$), Involvement ($M(SD) = 3.3(1.1)$), Spatial Presence ($M(SD) = 4.1(0.9)$), and General ($M(SD) = 4.9(0.8)$). Participants' immersion in the simulated SAMoD can be judged to be quite high. Participants compared the simulated AMoD journey to using public transportation systems such as buses or metros today (16 mentions; 6 women, 6 men).

Information Preferences

Overall, the qualitative data obtained in the study show that participants favored to have information about their co-passengers (15 mentions; 9 women, 6 men) over having no information (8 mentions; 2 women, 6 men). The most important reason for preferring the UI version with co-passenger information was security (22 mentions; 12 women, 9 men): *"I would have felt more secure with the display with the information and picture"* (P05), *"I felt so much more secure compared to the other version."* (P22). Participants considered the information as more pleasant (9 mentions; 5 women, 4 men) in terms of being connected to others: *"when the person comes in and you have a little info about them, I thought that was pleasant. You could also – in case something happens –*

address them by name or, yes, it is more pleasant than the anonymous [version]" (P04). Other advantages of having knowledge about fellow passengers were that participants considered it to be more interesting (4 mentions; 1 women, 2 men) and humane (3 mentions; 1 women, 1 men). Participants who preferred having information about other passengers where also willing to share these information about themselves. In line with this finding, the most important reason for preferring the UI version without information was privacy (17 mentions; 4 women, 6 men): *"My first thought was 'Oh no, people will know my name'. I don't like that at all."* (P12). Other participants regarded this information as not important (4 mentions; 1 woman, 3 men). Displaying passengers' details was even seen as insecure (3 mentions; 2 men) or untrustworthy (2 mentions; 1 man) as these details could potentially harm people. One participant expressed worries about the security of our young faux passenger ("Anna") as he elaborated: *"Well, at night you're just a bit more insecure, for example, when drunk, young people hop on. So [my worries] were also related to Anna, because people might think 'Oh, here comes Anna now, maybe we can hit on her or something.' That would be quite insecure for her then"* (P05). Although of our 24 participants, 15 expressed they would prefer the UI version with information, we would like to point out that this was not a clear decision every time. One participant even was unable to decide which version they preferred. Most participants found pros, as well as cons, for both versions and were weighing these off until finally making a decision. While this reflects how security and privacy are antagonists, the appropriateness of the variants was considered to be highly context-dependent, as outlined in the following paragraphs.

Day vs. Night

Generally speaking, our qualitative data confirms the difference the time of the day makes for sharing rides in SAVs with strangers as their number of mentions is higher (35 mentions; 9 women, 7 men) than statements that do not emphasize this importance (9 mentions; 2 women, 7 men). In this context, participants stated time-related concerns like *"during the night one is generally more careful and feels vulnerable* (P09). Several women (17 mentions; 9 women) expressed concerns when sharing rides with unknown men and said they would favor sharing a vehicle with other women at night over mixed vehicles. For instance, (P03) explains *"well, especially in the dark. During the day is not that tragic, but in the dark, I don't want to share a ride with a man or get off the vehicle with him."*. Interestingly, some of the men in our study conveyed similar feelings towards sharing rides with other men (8 mentions; 7 men) – particularly at night: *"Because it was Brigitte who got on at the first stop and then at the second stop it was a gentleman. That indeed made a difference to me"* (P11). As a reason they stated to feel more secure as a statement like, i.e., *"men tend to be more aggressive"* (P05) indicates. Participants also made clear that they would not need the displayed information on co-passenger during the day, but would prefer to have the information during the night: *"Especially at night it was more pleasant for me and more important. [...]. The fact that I was registered, for example, the [man/woman], as well. Yes, that was much more important for me at night than during the day"* (P22).

Type of Information

We asked participants which type of information they considered to be the most important one/s. The fellow passenger's profile picture was regarded to make all the difference (23 mentions; 7 women, 8 men) since it gives *"an impression of the person that is going to get on the vehicle at a glance"* (P16). In this regard the photo seem to give participants a feeling of control over the situation while the other information provided was rather a *"nice-to-have"* (P23). Knowing beforehand who would enter the vehicle also conveys security: *"Yes, I mean, I saw the picture and it looks nice and I actually had less fear"* (P03). The co-passenger's gender was essential, as well (13 mentions; 5 women, 2 men), followed by their age (10 mentions; 6 women, 2 men), the name (8 mentions; 3 women, 2 men), and the respective destination of the co-passengers (7 mentions; 2 women, 2 men). Most of the participants in our study stated that the information the system was offering was sufficient and emphasized how helpful it was to see the vehicle's route and its arrival time on the display. Some participants provided improvement suggestions such as getting information in case people with special needs, big luggage, or strollers would enter the vehicle, or whether seat belt use was compulsory.

6.3 Discussion

Overall, the results underline people’s openness towards SAMoD, which is in line with previous work [227]. Participants considered SAMoD to be useful and reported relatively high trust in the technology, intention to use, and positive experiences of the (simulated) SAMoD rides. However, participants also expressed concerns regarding security – especially with regard to night rides. Below, we discuss our findings in detail and situate them among previous work.

6.3.1 Night Trips Require Higher Levels of Information

In general, the SAMoD rides during the day were evaluated more positively than night rides. Participants consider the overall attractiveness of the SAMoD system higher and report more pleasant rides during the day. Rides without information about co-passengers were perceived as more pleasant than rides with such information. We hypothesize two reasons as sources of this findings: 1) participants are used to receiving no information about others when sharing a ride (as is the case in public transportation), and 2) people generally prefer rides during the daytime. This interpretation is comprehensively supported by our qualitative data and is in line with existing data from research in public transportation [25, 120, 194].

In contrast, rides with information provided by the in-vehicle UI were experienced to be significantly less risky compared to rides without information about co-passengers. Again, this is reflected in our qualitative data, with 21 participants underlining increased perceived security through the information. This can be taken as a general preference for information about co-passengers — particularly during the night and is in line with Ahmed et al. [1], who found that people are willing to provide information such as their gender, age, etc. to visually impaired persons in public spaces, if higher security assurances can be made. During the day, information about fellow passengers seems to have rather adverse effects (e.g., in terms of emotion). This, however, changes during the night, where it has, on the contrary, positive effects. Prior work underlines the importance of privacy particularly in public transportation [80]. Security and privacy are often antagonists in today’s public systems. This became evident in our study as participants mentioned privacy concerns when displaying personal information about other passengers, or themselves. During the interviews, participants weighted the pros and cons of having (no) information. Despite a preference for information during the night, this was not a clear outcome, which is also apparent in a higher dispersion of the Affect Grid assessments for the rides with information. While the information on co-passengers positively influenced security for some participants, there were also concerns that this information could have a negative effect exactly on security as strangers would know one's name and destination. To overcome the conflict between security and privacy, it needs to be investigated which information people feel comfortable sharing in order to increase perceived security.

6.3.2 Participants Prefer Sharing Rides with Women

While both men and women generally considered SAMoD systems useful, women rated them significantly more so and uttered a higher intention to use such services. We assume that finding is related to their (security) concerns in today's public transportation systems, especially considering night rides [194, 211, 227, 228]. In combination with qualitative data and the discovered interaction effect of passenger information and gender, this finding provides evidence that women seem to consider SAMoD systems to be more secure than 'classic' public transportation. Women and men alike explained in the interviews that they prefer sharing rides with women. This is in line with the findings of [139], who found people have higher refusal rates towards men as co-passengers. On the other hand, Polydoropoulou et. al [196] found different preferences of passengers for sharing with women/men between countries and cultures and that the number of fellow travelers further influences those preferences. In our study, we focused on rides with only one co-passenger as we expected this constellation would have the biggest effect on security. However, our results and the results from previous work [139, 196] underline once more the complexity of the topic.

6.3.3 Balancing Security and Privacy as a Design Challenge

In terms of overall SAMoD system design, there is most likely no 'one-fits-all' solution [169]. As, e.g., passengers' security needs are higher during the night, our data points toward flexible solutions for different times of the day

Based on our results, we propose that UIs for ride-sharing should provide general information on the route, arrival time, subsequent stops, and further information and functionalities to increase passengers' (feeling of) security for night rides. Providing information on fellow travelers can serve as a suitable option to do so. In our study, having a photo of fellow travelers was considered the most important information unit and was beneficial for passengers' feeling of security, while information on age, name, and destination played a subordinate role. In the study, we chose portraits with neutral facial expressions. However, other expressions might induce different – positive or adverse – feelings, e.g., feelings of insecurity. Given that photos seem to provide passengers with (at least some feeling of) control over the situation, these might be used in booking apps or in-vehicle displays. Passengers could then look for an alternative vehicle, or leave the vehicle at the next stop if someone's photo would make them feel uncomfortable. The feeling of control has been shown to have a positive effect on psychological security in the context of public transport [99] and, based on our results, we hypothesize that displaying a photo fosters this control. However, given the disagreement among our participants and the aforementioned privacy issues, we suggest 1) not exposing sensible data about co-passengers during the ride and 2) to consider alternative approaches to address participants' concerns. In terms of (1), it might be beneficial to relocate the information retrieval about fellow passengers to another time and place, e.g., the booking phase. For instance, [139] compared private and shared options on a mobile booking app and found that people tend to rather opt for shared rides when having detailed information on their fellow travelers prior to booking. This could also serve as a means to increase (perceived) security. In terms of (2), Schuß et al. [226] propose a "buddy system" to address women's security needs (during the night) that takes advantage of the fact that other passengers can also provide security. Instead of seeing them as potentially harmful, their approach focuses instead on the fact of not being alone and feeling secure instead of the feeling of controlling the situation through information. The concept of "social passengering" [162] among passengers inside the same or different vehicles points to a similar direction and might be beneficial for the perceived security.

6.3.4 Limitations

SAMoD is still a relatively 'theoretical' subject [193] with real-life applications remaining missing. Therefore, we let our participants experience a SAMoD system in a simulated environment. While participants report high presence perception and immersion, external validity is impaired due to the lab-based setup. As we were weighing off the negative side effects that come with lab studies, we opted for the simulated environment over conducting, e.g., a WoOz study in real traffic conditions, to compare the study conditions while ensuring high internal validity and high controllability.

As mentioned in section 6.1, we decided to simulate the presence of other passengers in a shared ride only virtually with sounds and display visualizations. While this was in line with the recommendation of [85] and facilitated the study's conformity with applicable hygiene regulations during the COVID-19 pandemic, the chosen representation of a shared ride's social context is limited. On the other, considering our study design with multiple measurements during a test ride, the physical presence of another person might have affected participants' assessment of the information and consequently the study's reliability. Furthermore, we did not intent to focus on the inherent social factors or mutual relationships (that definitely play an essential role in the context of shared mobility), but focused on the provided information. Nevertheless, this should be considered when conducting further studies on SAMoD.

Taking into account the large and diverse population of future SAMoD users, our study has been conducted with a small sample and, although having placed value on a broad spectrum of people (gender-balanced, different age groups, different cultural backgrounds, different education levels), it covers only a part of the variety of potential users. According to the STAI inventory, our participants had relatively low levels of trait anxiety. Since this trait likely has an effect of risk and security evaluation, generalizability is limited.

Furthermore, the study was conducted during the COVID-19 pandemic. We applied precautions like distancing and hygiene measures and followed the regulations of local and national authorities. While we consider the pandemic's effect on the study conducted to be minor, it might have affected the sample composition as, e.g., only people with medium fear and anxiety have signed up for the study. It would be interesting to repeat this study with people that show higher levels of trait anxiety as this trait influences the evaluation of risk and security of situations, and we hypothesize that these people could have evaluated the presented prototype in a more positive way.

The selection of the displayed information on co-passengers covers only a part of the potential variety and might have fostered stereotypes. We derived the solution with information about co-passengers based on existing research findings [227, 139] and aimed to evaluate whether the availability positively influences security, UX, trust, and acceptance of SAMoD passengers. By no means we intended to manifest potential stereotypes or the exclusion of people through our selection. However, we want to point out that the selection likely affects the results (e.g., people might refuse rides with others due to their "look"). We are aware that the gender and age of other passengers is a limited view. Other factors, such as race, appearance, or the supposedly associated social statuses definitely play a role in people's assumptions about other people. However, we did not include more personal characteristics to 1) not confound too many different independent variables in the display variants and 2) we wanted to draw a clear line between evaluating the information about other passengers and participants' potential biases about, e.g., other cultures, as we aimed for the former.

6.3.5 Future Work

Passengers' information demands in SAMoD systems are a highly complex and context-dependent issue requiring more research, especially on how to overcome the conflict between security and privacy by design.

Based on our results, we suggest extending the conduct of context-based empirical studies investigating factors like daytime and fellow passengers in SAMoD systems along the whole travel journey. Since, e.g., security issues are relevant for the booking, the ride itself, and on-/off-boarding [227]. While our study focused on the ride itself, further (empirical) studies should also consider the booking phase and the off-boarding when investigating the effect of co-travelers and time of day on passengers' need for information and controls. Here, additional information and safety measures (e.g., emergency/support button) might support passengers' feeling of control and security. To yield results with high external validity, future studies might include more contextual factors such as the (physical) presence of various and multiple other people in SAMoD rides during different situations. E.g., actors could be used to mimic specific situations [85].

It would also be interesting to repeat this study in different cultural contexts, as we conducted our study in Germany, where security in public transportation offers high levels of security [99]. However, we assume that conducting similar studies in countries, such as India or Latin American countries, where public transportation is more difficult to access – especially for women [120] – might yield different results. The applied simulation environment presents a context-based prototyping approach that can be used, e.g., to replicate this or similar studies in other countries and investigate potential cultural differences regarding passengers' (information) requirements. Future work might also consider the potential impact of culture and race as an independent variable in the information display. This could result in an exploration of people's explicit and implicit biases based on given prior knowledge.

We used the front of the vehicle as the output location of the information, as these are common modalities [126]. Future concepts might also investigate whether (the combination with) other modalities, such as tactile, influence the perception of the presented information and the feeling of security.

6.4 Conclusion

In this chapter, we report on a simulator user study ($N = 24$) investigating the effects of time of day and information provided about fellow travelers on SAMoD passengers' UX, acceptance, feeling of security, and emotions in shared automated rides. While the evaluated SAMoD system received excellent assessments of hedonic and pragmatic UX, trust, and acceptance, participants emphasized security concerns – mainly when using SAMoD at night. Furthermore, both women and men preferred sharing rides with women over sharing rides with men as co-passengers during the night, whereas, during the day, this information negatively affected participants' evaluation of the SAMoD system. Associated risks were experienced lower when participants were provided with information about their co-passengers. Most participants generally preferred having information on co-passengers, with photos of fellow travelers considered to be the most important information element. However, our results yield ambiguities since providing personal information also triggered privacy concerns among participants. This can be taken as an illustration of the complexity of psychological security and its context dependency. Building upon these findings, providing UIs with information on fellow passengers can support SAMoD passengers' feeling of security in shared rides and potentially improve UX and overall acceptance. However, due to privacy concerns and associated risks, the timing and placement of the information need to be questioned. It might be beneficial to provide this information during the booking phase, but not within the vehicle. Future work should consider the complete travel journey of SAMoD, foster the inclusion of contextual factors, and investigate how the provision of additional information and safety measures (e.g., emergency and support features) can increase passengers' feeling of control and security.

6.5 Chapter Summary

Security concerns are a critical barrier to the public adoption of SAMoD systems since passengers have to share rides with strangers without a human driver on board. These concerns seem especially prevalent for night rides.

Motivated by related work that points out that information about fellow travelers might mitigate passengers' concerns, this chapter investigated the role of information about co-passengers in an exploratory within-subjects user study ($N = 24$). Participants experienced two day rides and two night rides in a simulated shared AV with varying personal information about co-passengers displayed on an in-vehicle passenger information HMI.

The results of the mixed-method study indicated that passengers demand more information about co-passengers during night trips. Information during night rides positively affected the overall attractiveness of the SAMoD system and resulted in more pleasant ride experiences. In contrast, during the day, participants deemed the information unnecessary. Generally, rides with information on co-passengers were experienced significantly less risky than rides without. Regarding the information provided by in-car interfaces, photos of fellow travelers appear to make the difference, while age and name are of subordinate importance. In terms of co-passenger details, both women and men preferred sharing rides with female co-passengers over sharing rides with male co-passengers. However, participants also raised privacy concerns regarding the display of personal information. Therefore, we concluded that future researchers and practitioners should balance people's security and information demands with privacy concerns.

For the simulation, we used an adaption of the immersive video-based prototyping approach presented in Chapters 3 and 5 with TVs and in combination with the tent-based mock-up by Schuß et al. [227]. In addition, the static mock-up increased participants' spatial immersion in the interior space of a simulated shared AV. Regarding the social context simulation, we opted for a virtual variant due to economic reasons, the prevalent COVID-19 pandemic at the time of the study's conduct, and based on the findings of Chapter 3. However, we acknowledge that this is, in particular regarding the study's design and research question, up for discussion. Since we specifically aimed to assess information related to the social context, actor involvement might have yielded higher external validity and clearer results. However, those would have depended very much on the characteristics of the particular actor. With the virtual simulation of various co-passengers, we were able to balance this in terms of demographics. Yet, we could only include some archetypes but could not cover the variety of potential passengers.

Through the investigation of Q6.1, this chapter contributes to our general reserach question **RQ1** with findings on user requirements and derived HMI design recommendations based on an empirical study on SAMoD passengers' information demands, particularly in terms of information on co-passengers during the day and night rides. Regarding **RQ2**, this chapter demonstrated how immersive video-based simulation could be enhanced using a static interior and exterior AV mock-up. Building on the concepts of Chapter 3, it also exhibited more detailed applications of virtual social context simulation.

7

Conclusions

This chapter starts with a summary of the dissertation (Section 7.1) and an overview of its contributions (Section 7.2). We then provide answers to our two general research questions on HMI design (RQ1) and prototyping (RQ2) of (in-vehicle) human-AV interactions (Section 7.3) and close this thesis by outlining potentials for future work (Section 7.4).

7.1 Summary

Motivated by the commencing introduction of AVs (i.e., vehicles with driving capabilities of SAE level 4 or 5 [210, 278]) to the public, this doctoral research investigated context-based design and prototyping approaches for (in-vehicle) human-AV interactions to counteract acceptance hurdles and facilitate adoption from early development phases. Therefore, we adopted a research in and through design approach [59].

Chapter 1 introduced the topic and methodology and motivated the research questions investigated within this doctoral research. After a glance at the history, future, and technical background of vehicle automation and corresponding concept scenarios, we dove into the topic of human-AV interaction. Since no driver is required to control the vehicle, AVs face general skepticism and public concerns that need to be overcome for

THIS CHAPTER IS BASED ON THE FOLLOWING PREVIOUS PUBLICATIONS. PARTS OF THIS CHAPTER WERE PUBLISHED PREVIOUSLY AS PART OF THEM.

[85] **Lukas A. Flohr**, Dominik Janetzko, Dieter P. Wallach, Sebastian C. Scholz, and Antonio Krüger. 2020. Context-Based Interface Prototyping and Evaluation for (Shared) Autonomous Vehicles Using a Lightweight Immersive Video-Based Simulator. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1379–1390. `https://doi.org/10.1145/3357236.3395468`

[86] **Lukas A. Flohr**, Sofie Kalinke, Antonio Krüger, and Dieter P. Wallach. 2021. Chat or Tap? – Comparing Chatbots with 'Classic' Graphical User Interfaces for Mobile Interaction with Autonomous Mobility-on-Demand Systems. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. `https://doi.org/10.1145/3447526.3472036`

[87] **Lukas A. Flohr***, Martina Schuß*, Dieter P. Wallach, Antonio Krüger, and Andreas Riener. 2023. Designing for Passengers' Information Needs on Fellow Travelers: A Comparison of Day and Night Rides in Shared Automated Vehicles. *arXiv* 2308.02616 (aug 2023), 22 pages. `https://doi.org/10.48550/arXiv.2308.02616`.
**Note: both authors contributed equally to this paper.*

[88] **Lukas A. Flohr**, Joseph S. Valiyaveettil, Antonio Krüger, and Dieter P. Wallach. 2023. Prototyping Autonomous Vehicle Windshields with AR and Real-Time Object Detection Visualization: An On-Road Wizard-of-Oz Study. *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. `https://doi.org/10.1145/3563657.3596051`

[89] **Lukas A. Flohr** and Dieter P. Wallach. 2023. The Value of Context-Based Interface Prototyping for the Autonomous Vehicle Domain: A Method Overview. *Multimodal Technologies and Interaction 7*, 4 (2023), 1–17. `https://doi.org/10.3390/mti7010004`

[260] Dieter P. Wallach, **Lukas A. Flohr**, and Annika Kaltenhauser. 2020. Beyond the Buzzwords: On the Perspective of AI in UX and Vice Versa. In *Proceedings of the 1st International Conference on Artificial Intelligence in HCI, Held as Part of the 22nd International Conference on Human-Computer Interaction* (Copenhagen, Denmark) *(HCII '20)*. Springer International Publishing, 146–166. `https://doi.org/10.1007/978-3-030-50334-5_10`

successful adoption. While people are already getting used to the increasing number of vehicles equipped with SAE level 2, AVs are still under development. They are currently only available with limitations restricting use cases and evaluation scenarios. To investigate and develop human-AV interactions that can counteract future users' concerns, prototyping methods are required that enable us to take the unfamiliar context of driverless road traffic into account from early development phases.

With **Chapter 2**, we situated this research among related work and provided a comprehensive background on human-AV interaction and context-based prototyping. We furnished an overview of acceptance and human factors challenges that need addressing for the successful adoption of AVs and supplemented it with practical considerations for HMI design by also examining current state-of-the-art solutions. This analysis served as a basis for investigating our general research question **RQ1** on designing suitable human-AV interactions. The identified challenges are especially prevalent in the particular physical, social, and situational contexts of driverless rides in urban traffic. Thus, we figured out the importance of incorporating context into all HCD activities and argued – also as a starting point for answering **RQ2** (prototyping) – for applying context-based interface prototyping [85, 118, 117]. As a conceptual understanding of context, we referred to Dey and Abowd [65], who describe context in the field of HCI as any information that characterizes the situation of a person, place, or object relevant to the interaction. Such context can incorporate physical, social, cultural, or organizational factors [253]. Here, this doctoral research focused on the physical and social context. Contextualized prototyping can condense associated contextual aspects together with the prototype of a product, system, or service "into one single manifestation" [117, p. 210]. With references to Lim et al. [154] and ISO 9241 [125], we derived that prototypes are most valuable if they help to achieve the goals of an analysis, design, or evaluation activity (most) efficiently and straightforwardly. Therefore, asking Wh-questions such as what, why, and when is crucial to deciding how a prototype should manifest. After this theoretical part, we took an in-depth look at suitable methods for context-based prototyping of (in-vehicle) human-AV interactions, namely 1) static mock-up, 2) ride simulation including virtual and mixed reality as well as video-based approaches, 3) social context simulation including actors, enactment strategies, social items, and props, 4) WoOz setups, in which a human driver simulates an AV on a test track or in real traffic, and 5) experimental vehicle, i.e., an actual (automated) vehicle (with restrictions).

In **Chapters 3, 4, 5, and 6,** we presented our empirical research following a research in and through design approach. Chapters 3 and 4 introduced the straightforward and cost-effective context-based prototyping approaches we created and applied in our empirical research. The approaches primarily contribute to answering **RQ2**. In particular, we demonstrated an immersive video-based ride simulation, social context simulation with actors, and a video-based WoOz setup. Chapters 5 and 6 use variations of the immersive video-based approach presented in Chapter 3 in combination with static interior and exterior mock-ups to investigate HMI concepts for mobile interaction with shared AMoD systems (Chapter 5) and passengers' information demands during shared AV rides (Chapter 6). The findings and design recommendations derived in Chapters 3, 4, and 6 provide the basis for answering **RQ1**. The following sections provide some more details about each chapter.

In **Chapter 3**, we adapted the approaches of Kray et al. [143] and Gerber et al. [98] to create an immersive video-based simulator for context-based prototyping of (shared) AV rides. We used low-budget action cameras to capture real-world video and audio footage of urban traffic, which we displayed on three video projectors in an office space. The three projections created a CAVE-like environment and resembled the views out of the left, front, and right windows of the simulated vehicle. We evaluated the setup regarding immersion and suitability for contextualized AV prototyping with an expert study ($N = 9$) and a user study ($N = 31$). In the latter, we also simulated the social context of a shared AV ride by using the sounds of other passengers and an actor that physically joined the simulated ride. The conducted studies and the assessments of the setup revealed the general suitability and reliability of the method. However, the approach is limited regarding controllability and – like most simulators – in terms of (external) validity. Actor involvement positively affected participants' feeling of reality, observable in our single-item measurement and participants' qualitative feedback. As this was not confirmed by our standardized immersion measures (IPQ) and was accompanied by higher discomfort (motion sickness symptoms), we did not generally recommend actor usage – also considering the extra effort and costs. However, we pointed out that involving actors can make sense to simulate scenarios with a particular interest in factors related to the social context.

Inspired by previous works of Karjanto et al. [131] and Detjen et al. [64], **Chapter 4** presented a straightforward and flexible WoOz setup for on-road AV simulation in real traffic. The setup separated the front and back area of an electric minivan with a curtain mounted on the vehicle's ceiling and a 43" TV mounted on the front seats' headrests. The TV displayed a live camera stream of a webcam mounted on the minivan's windshield and provided passengers with a video-based replacement of the front view. In the conducted study, we used the WoOz setup to prototype and evaluate a futuristic AR windshield display that provided computer-vision-based information on detected objects (e.g., other vehicles, pedestrians, traffic lights) calculated in real-time on an Nvidia Jetson Nano (question Q4.3). To investigate the potential of information transparency (questions Q4.1 and Q4.2), we conducted a user study with $N = 30$ participants assessing three HMI variants: 1) no information on detected objects (baseline), 2) abstract status bar with counts of the detected object classes, and 3) AR overlays. System feedback on detected objects, especially through AR-based visualization, increased system understandability and hedonic quality. However, participants pointed out that the information display could also distract, irritate, or annoy, and thus demanded customization options. Since participants missed travel information such as current location, upcoming stops, and estimated arrival times, we figured it is essential to investigate the transparent information concept considering the interplay with other HMIs. With 73 % of the sample believing the WoOz deception, we see the setup as a suitable method for context-based interface prototyping, particularly regarding studies in real traffic and (AR-based) HMI concepts using real-time information.

In **Chapter 5**, we compared two concepts for mobile interaction with AMoD systems – 'classic' GUIs and chatbots (i.e., text-based conversational UIs) – to investigate question Q5.1. We presented the results of two expert studies ($n_{\text{GUI}} = 6$; $n_{\text{Chatbot}} = 5$) and a between-subjects simulator study ($N = 34$; $n_{\text{GUI}} = 17$; $n_{\text{Chatbot}} = 17$). The simulator study used a TV-based adaptation of the immersive-video-based setup presented in Chapter 3 to simulate a complete AMoD user journey considering interactions before, during, and after the ride. Comparable results of participants' immersion hint at the successful replication of the cost-effective setup. Both tested concepts received good assessments in terms of acceptance and UX. The GUI seemed superior with regard to attractiveness and user satisfaction assessments. However, in scenarios with involving changes of plans, participants indicated a higher intention to use the chatbot. We summarized the findings from the mixed-method study and set them in the context of related work to provide an overview of the pros and cons of using the two concepts. Based on the results, we recommended using GUIs as a basis for mobile interaction with AMoD but to increase user assistance and guidance, particularly in scenarios with changes of plans, by providing a conversational UI component as a supplement.

Since related work suggested providing information on fellow travelers to mitigate AMoD passengers' security concerns, we conducted an exploratory within-subjects user study with $N = 24$ participants on which we reported in **Chapter 6**. The study used a TV- and immersive-video-based simulator similar to the one used in Chapter 5 in combination with the spatial mock-up applied by Schuß et al. [228] to enhance spatial immersion. For the social context simulation, we opted for a virtual representation of fellow travelers through auditory features (i.e., sounds of other passengers entering and leaving the vehicle) and visual display on the passenger information display. This supported aligning the study with COVID-19 regulations at the time of the study conduct and meeting budgetary restrictions. We acknowledged that regarding the matter of subject, actor involvement (such as applied in Chapter 3) may have been beneficial in terms of participants' immersion and the study's external validity. We let participants experience two rides during the daytime and two during the nighttime with varying amounts of information on fellow passengers to investigate our research question Q6.1. Results revealed a positive effect of information provision (particularly images of fellow travelers) during night rides on the attractiveness of the shared AMoD system and passengers' ride experience. However, during the daytime, the information was assessed as unnecessary. Both women and men preferred to share rides with female co-passengers. While in general, rides with information on co-passengers were assessed as less risky than rides without, participants raised privacy concerns. We concluded that balancing these apparent antagonists (security vs. privacy) poses future design challenges for successful AMoD systems.

Below, we channel our work and learnings into contributions to answer our general research questions on the design and prototyping of human-AV interactions.

7.2 Contributions

Despite technological hurdles, AVs face significant acceptance and UX challenges that need to be overcome for their successful adoption. To address those, we identified the necessity to design suitable HMIs for trustful human-AV interaction, which resulted in the formulation of RQ1. Furthermore, since the availability of AVs is still limited for researchers and practitioners and the consequent access to the context of driverless rides, we investigated methods for early and straightforward prototyping and evaluation incorporating the dynamic physical and social context of (in-vehicle) human-AV interactions with RQ2. Tables 7.1 and 7.2 summarize our contributions to our two general reserach questions and cluster them according to their contribution type.

Table 7.1 – Overview of contributions to RQ1: How might we design suitable HMIs for (internal) human-AV interaction that counteract acceptance challenges?

Type	Contribution	Chap. [Ref]
Theoretical	Qualitative semi-systematic literature review [236] of acceptance and human-factors challenges that need consideration in the design of HMIs for human-AV interactions.	2
	Literature-based overview of prospective passengers' information needs and applicable HMI frameworks and concepts.	2
Practical, Empirical	Design recommendations for mobile interaction with AMoD systems and comparison of 'classic' GUIs and chatbots (text-based conversational HMIs) through literature review, expert studies ($N_{\text{expGUI}} = 6$, $N_{\text{expChatbot}} = 5$) and a comparative user study with $N_5 = 34$ participants.	5 [86]
	Design recommendations for shared AMoD systems based on an empirical study ($N_3 = 24$) on passengers' information demands, particularly in terms of information on co-passengers during day and night rides.	6 [87]
	Prototyping and design recommendations for novel/futuristic AV windshield concepts using real-time information and AR-based visualizations based on an empirical WoOz study conducted on real urban roads ($N_4 = 30$).	4 [88]

Table 7.2 – Overview of contributions to RQ2: How might we prototype and evaluate (internal) human-AV interactions early considering their dynamic context?

Type	Contribution	Chap. [Ref]
Theoretical	Qualitative semi-systematic literature review [236] of related work carving out the value and impact of prototyping the physical and social context for designing human-AV interactions.	2 [89]
	Method overview for context-based AV interface prototyping and recommendations for their application.	2, 7 [89]
Methodological	Immersive video-based approach for simple, flexible, and cost-effective prototyping of in-vehicle human-AV interactions; initial evidence of the method's suitability from an expert study ($N_{\text{expSim}} = 9$) and three user studies ($N_3 = 31$, $N_5 = 34$, $N_6 = 24$).	3, 5, 6 [85, 86, 87]
	Demonstration of how immersive video-based simulation can be combined with static interior and exterior AV mock-ups to enhance spatial immersion.	6 [87]
	Video-based WoOz approach for on-road simulation of AV rides and AV windshield interfaces with AR and real-time (object detection) information that can easily be recreated and transferred to other vehicles.	4 [88]
Practical, Empirical	Investigation of the (practical) value of actor engagement for social context simulation of shared AV rides, i.e., the physical presence of other passengers ($N_3 = 31$).	3 [85]
	Using sounds and information display as a virtual simulation of other passengers in a simulated shared AV ($N_3 = 31$, $N_5 = 34$, $N_6 = 24$).	3, 5, 6 [85, 86, 87]
	Recommendations for WoOz study planning and conduct based on an empirical user study in real traffic ($N_4 = 30$).	4 [88]

7.3 Answers to the Research Questions

Based on our learnings and contributions (Section 7.2), we derive the following answers to RQ1 and RQ2. Those encompass practical design and prototyping recommendations as well as the initial basis for a methodological (decision) framework for context-based prototyping of human-AV interactions.

7.3.1 RQ1. How might we design suitable HMIs for (internal) human-AV interaction that counteract acceptance challenges?

Riding in a (shared) AV without a human driver might feel awkward for passengers since they are fully exposed to an autonomous, AI-powered system's perceptions, decisions, and actions. HMIs (or UIs) are their only way of communicating and interacting with driverless vehicles. Thus, HMIs must compensate for the resulting service and information gap and counteract users' concerns to facilitate acceptance and the technology's adoption. Below, we structure our learnings into four recommendations.

Consider human factors and acceptance challenges from early development phases with thorough application of an iterative HCD process.

To design appropriate HMIs for human-AV interaction – and to provide answers and solutions in terms of RQ1 – we need a precise understanding of people, context, and the (autonomous) systems [149], as well as of associated problems and concerns. Therefore, it is essential to understand human factors and acceptance challenges (such as trust issues, passengers' UX and situation awareness, and their concerns regarding safety, security, and privacy) and consider them when designing HMIs for AVs (Chapter 2). We argue that applying the HCD process [125] offers a suitable methodological framework to address the mentioned challenges and consider prospective users' needs, such as information demands and interaction preferences from early phases. Frameworks such as Collaborative UX Design [242] provide suitable methods to address inter- and intraindividual needs and perspectives and tackle complex problems. In the sense of HCD, they facilitate incorporating actual and prospective requirements of users and stakeholders. From our perspective, considering contextual factors is vital in all phases of the HCD process. We address this aspect in more detail through context-based interface prototyping regarding RQ2 (Section 7.3.2).

Provide transparent system feedback to increase understandability and trust, but do not overload while considering the whole picture.

In the sense of HCAI [204] and XAI [105], HMIs can and should help humans to understand AI-based AVs. Among other things, this is also part of the guidelines for human-AI interaction postulated by Amershi et al. [6]. To achieve such an understanding, this may require HMIs to provide appropriate feedback [183], i.e., to provide explanations and to make the process within the intelligent technical systems more transparent to the user. We investigated the provision of real-time feedback on AV's information processing and found potential to increase the system's understandability and users' hedonic UX (Chapter 4; Q4.1). On the other hand, the location, time, and (visual) design of such information need to be carefully chosen and crafted, considering that our findings indicate that providing such information may also have adverse effects and, e.g., annoy, irritate, or distract users (Q4.2). Consequently, and as Oliveira et al. [185] also point out, the amount of information provided matters. Not all users want such transparent system feedback (at all times). We, thus, concluded making information displays configurable for passengers as a suitable approach. Our WoOz study (Chapter 4), furthermore, highlighted that it is crucial to design and investigate the interplay of new HMI concepts (e.g., AR-based system feedback on object detection) from a holistic perspective that includes 'basic' information passengers would expect.

Provide users with travel information and functions they are already familiar with in combination with specific needs and demands for (shared) AV rides.

With regard to information demands, related work pointed out that prospective AV passengers demand and expect that AVs provide similar information as they are used to by (non-autonomous) private and public modes of transport (e.g., information on current status, planned route, and upcoming events [24, 159, 193]) – at least in the early phases of market launch when the technology is still new to users. This is also confirmed by our simulator studies (Chapters 3, 5, 6) and particularly by our WoOz study (Chapter 4), where many participants missed such information during the on-road AV ride simulation.

Similarly, this also applies to familiar devices and interaction technologies. Users seem, e.g., to prefer familiar concepts such as smartphone apps, touchscreens and physical control buttons in the vehicle over new modalities, such as gesture communication [24]. More generally, established modalities such as visual, auditory, and haptic

[126] are most suitable for basic in-vehicle communication. However, multi-modal interaction design is important for in-vehicle communication [126].

Since in shared AVs and AMoD, rides are shared with other passengers, and travel information can be a private matter [30], visual in- and output on personal devices seem superior to 'more public' interaction modalities such as auditory. As a result, we recommend combining public in-vehicle information displays providing general information on the ride and status with personalized travel information on mobile devices. In Chapter 5, we investigated 'classic' GUIs and chatbots (i.e., text-based conversational UIs) for mobile human-AV/AMoD interaction (Q5.1). Generally, both concepts can serve users as travel companions throughout the whole (AMoD) travel journey (before, during, and after the ride). Based on our empirical findings, we recommend using 'classic' GUIs as a basis for mobile interaction since participants rated them significantly more attractive and satisfying. Nevertheless, we also found that chatbots can support users in demanding situations (e.g., in scenarios involving a change of plans), so we also recommend integrating (alternative/additional) conversational features. Mobile travel companions (with conversational elements) could also (partly) compensate for missing human chaperones in autonomous rides.

Carefully weigh passengers' security, safety, and privacy needs.

Particularly in shared AVs and AMoD, HMIs also need to compensate for the absence of a human driver to account for passengers' safety and security concerns (e.g., fear of assaults). Besides providing transparent system information, this could require the provision of safeguards and information on fellow passengers. Our results reveal that prospective passengers have higher information demands during AV rides at night (Chapter 6; Q6.1). Here, passengers feel more secure when they receive information on their co-travelers (e.g., name, age, gender, picture, and destination). However, study participants displayed ambiguous feelings about receiving such information from others and providing information about themselves. Particularly the latter triggered privacy concerns. Our results from Chapter 6 highlight the tension between passengers' safety and security needs and their privacy demands. Future system designs need to address this conflict by investigating and defining which information is required and beneficial to enhance perceived security, on the one hand, and, on the other hand, people are also comfortable sharing.

7.3.2 RQ2. How might we prototype and evaluate (internal) human-AV interactions early, considering their dynamic context?

Prototyping is crucial in the domain of (shared) AVs and AMoD, especially in the early design and development phases. Among other considerations, this is mainly because the availability of actual (driverless) AVs – and consequently the access to the context of driverless rides – is still limited. Consequently, designing human-AV interactions means designing future and futuristic mobility systems. This implies that many HMI (or, more general, service/system/product) design decisions are based on assumptions that need to be analyzed, tested, and validated considering the context of use. Prototyping can support such design and research activities and facilitate applying an HCD process. Based on our research, empirical findings, and practical learnings, we postulate the following three general recommendations as answers to RQ2. The recommendations may serve as a basic (decision) framework for prototyping human-AV interactions in early design phases.

Consider and question the what, why, how, for whom, and where something is prototyped and the resources required for an effective and efficient design process.

Depending on the question and context at hand, various approaches and methods are suitable to prototype human-AV interactions. Based on our theoretical and practical research, we can conclude that there is no such thing as the perfect prototype or the perfect tool. However, we see that there are particular factors that highly impact the success and efficiency of prototyping activities such as the timing and the resources used (see Section 2.3). In Section 2.3.4, we rendered Buxton's [38] perspective that "everything is best for something but worst for something else" and that it is essential to know "what, for what, when, for whom, where, and [. . .] why" something is prototyped. We furthermore introduced Dodge's [70] equation on factors affecting a prototype's impact on design.

Although Buxton [38] did not explicitly refer to prototyping methods but to the choice of input devices, we want to note the close relation and the overlap to Dodge's [70] impact of prototyping equation (Equation (2.1)). In fact, we propose to extend Dodge's [70] equation with the "missing" aspects of *for whom* and *where*. From a human-centered design perspective, for whom (i.e., for which audience or which particular stakeholders) we prototype is an essential consideration for creating prototypes. Similarly, *where* can be regarded as the representation of the product's, system's, or service's physical context but also the prototype's own physical location. Where we create, implement, or test a prototype may significantly affect its manifestation and consecutive results. As a consequence, we propose to combine Dodge's [70] equation with Buxton's *for whom* and *where* [38] as laid out in Equation (7.1).

Additionally, unlike our previous publication discussing this equation [89], we generalize the *time spent* to the *required resources*. We take into account that besides the timely dimension, a prototype's impact on the design process can also depend on the required amount of other (e.g., economic, monetary, physical, or even social) efforts and expenses.

$$\frac{\text{What} \times \text{Why} \times \text{How} \times \textit{For Whom} \times \textit{Where}}{\textit{Required Resources}} \times \text{When} = \text{Impact on Design} \tag{7.1}$$

In line with Equation (7.1), Warfel [262] emphasizes that it is important to understand the audience (i.e., for whom) and intent (why) and to prototype only *what* you need. Furthermore, Warfel [262] points out that prototyping is a generative and iterative process which leads to the recommendation to prototype "early and often" ([262], p. 95). Table 7.3 provides a collection of example variables for Equation (7.1)'s factors.

Table 7.3 – General and AV-domain-specific example variables for the factors affecting a prototype's impact on the design (process).

Factor	Example Variables
What?	HMI/UI: modality (e.g., visual, auditory, tactile), type (e.g., iHMI, aHMI, eHMI), device (e.g., info display, light display, smartphone), prototype fidelity (e.g., low, high); Context: physical context (e.g., AV interior, AV ride through an urban environment), social context (e.g., interaction with co-passengers); situational aspects (e.g., time of day, weather);
Why?	HCD activities: understanding, specifying, producing, evaluating; exploration; (public) demonstration; validation;
How?	Static mock-up; ride simulation; social context simulation; wizard-of-oz; experimental vehicle;
For whom?	User; customer; manager; citizen; designer; developer; other stakeholder;
Where?	Lab; (restricted) test area; real (urban) roads;
When?	Early/late in the design process; ongoing development/optimization;
Required resources	Time; money; material; equipment; (specific) skills;

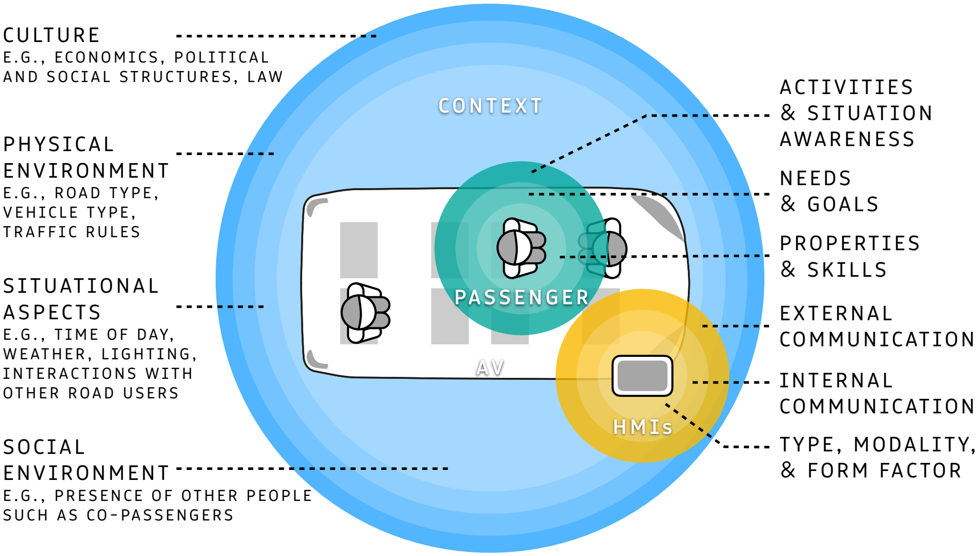


Figure 7.1 – Influencing factors on (in-vehicle) human-AV interaction that should be considered in design and prototyping activities. The layer illustration is based on original diagrams of Savio and Braiterman [215] on mobile interaction and Hoggenmüller [117] on interactions with urban robots. The illustration, furthermore, refers to concepts and frameworks discussed in Chapter 2.

Apply context-based interface prototyping to incorporate influencing factors early.

Since the availability to experience driverless rides' physical and social context is still limited – especially in the early design stages, we argue for applying context-based interface prototyping [85, 88, 118]. Context-based prototyping situates the envisioned HMI/UI – or, more generally, the envisioned product, system, or service – in the context of use (Figure 2.14, Figure 7.1). It increases users' and stakeholders' immersion and the scope of designers and researchers [118]. Consequently, it enables considering the dynamic context of driverless AV rides (e.g., through complex urban environments) from early development phases. Thus, context-based interface prototyping provides a suitable basis for efficiently exploring and evaluating (in-vehicle) human-AV interactions. Figure 7.1 illustrates influencing factors during (shared) AV rides and their interrelations resulting from context (i.e., culture, physical and social environment, and situational aspects), passenger (i.e., AV users' activities and situation awareness, individual needs and goals, and properties and skills), and HMIs (i.e., HMI type, modality, and form factor, internal and external communication).

Table 7.4 – Our assessment of the methods' suitability for supporting the four key activities of the HCD process (understanding, specifying, producing, and evaluating) and regarding their level of immersion/realism, (external) validity, reliability, cost-efficiency, and options to combine them. Simple rating scale: ■□□ = low, ■■□ = medium, ■■■ = high. The methods' actual suitability/score, however, may vary depending on influencing factors and corresponding variables (Table 7.3).

Method	HCD: Understanding	HCD: Specifying	HCD: Producing	HCD: Evaluating	Immersion/Realism	Validity	Reliability	Cost-efficiency	Combinable with
Static mock-up (1)	■■□	■■□	■□□	■■□	■□□	■□□	■■■	■■■	(2), (3)
Ride simulation (2)	■■□	■■□	■■□	■■■	■■□	■■□	■■■	■■■	(1), (3)
Social context simulation (3)	■■■	■□□	■■□	■■□	■■□	■■□	■■□	■■□	(1), (2), (4), (5)
Wizard-of-Oz (4)	■■■	■■□	■■□	■■■	■■□	■■■	■□□	■■□	(3)
Experimental vehicle (5)	■■■	■■□	■■□	■■■	■■■	■■■	■□□	■□□	(3)

Select and combine prototyping methods fitting your needs and design stage.

Generally, several approaches, methods, and tools for context-based prototyping are available. Opting for one or a combination should depend on previous considerations of factors affecting a prototype's impact on the design and factors influencing the interaction between humans and AVs (Table 7.3, Figure 7.1).

In Section 2.4, we provided an overview of applicable context-based prototyping methods for contextualized exploration and evaluation of human-AV interactions and discussed their respective challenges. Based on our learnings from our prototyping activities and user studies (Chapters 3, 4, 5, and 6) and from our literature reviews (Chapter 2), we conclude practical recommendations for their application and summarize them in Table 7.5. The table is an extension of our previously published list [89] with further recommendations and details derived from the integrated review of the works conducted in this doctoral research.

Table 7.5 – Our recommendations for the application of the discussed context-based prototyping methods in human-AV interaction design.

Method	Recommendations
Static mock-up	Use static mock-ups for spatial prototyping of AV interiors and exteriors in lab environments (Chapter 6); Combine with other methods to increase immersion, e.g., ride simulation and social context simulation (Chapters 3, 5, and 6).
Ride simulation	Use (immersive) real-world videos for (passive) high-fidelity simulation (Chapters 3, 5, and 6) in lab environments. Since real-world representations can increase participants' familiarity with the context [118], consider the resulting possible desirable (but in some scenarios also possibly undesirable) side effects thereof before choosing video-based approaches; Use (CGI-based) VR and MR to prototype interactive environments, i.e., contextual representations that need to respond to human input (e.g., to investigate what happens when the user requests an emergency stop during an AV ride); Choose an appropriate duration for the simulated rides reserving enough time for participants to become immersed in the simulation but not bored. Based on our experiences, we propose 8 to 10 min per ride as a rule of thumb for HMI variant comparison in sequential rides; Consider combining immersive video and VR/MR [98] to get the best of both worlds if required resources are available; Use the checklist provided by Hock et al. [116] to design valid simulator studies and address issues such as simulator sickness.
Social context simulation	Use social context simulation to consider the effects of other people present in particular situations and the users' relationship with them [143], e.g., co-passengers joining a shared ride at night (Chapter 6); Use sound as a baseline for social context simulation. For instance, to simulate a shared ride, enhance the audio from the physical environment with noises of people entering or leaving the vehicle (Chapter 3); If social aspects are a crucial facet of the study, consider using actors (Chapter 3) and enactment [228] to increase immersion and validity.
Wizard-of-Oz	Use WoOz to explore user behavior and investigate new concepts in complex real-world environments, e.g., a futuristic real-time AR-based windshield HMI (Chapter 4); Use fitting cover stories ([64], Chapter 4) to introduce and maintain the WoOz deception of participants; Support the story with a consistent 'AV-like' driving style (e.g., "like a professional limo driver" [11]) and fitting hardware (Chapter 4); Shift participants' focus away from the WoOz setup to other points of interest, such as the HMI concept that needs to be assessed (Chapter 4).
Experimental vehicle	Transparently inform participants about the vehicles' capabilities (if not in conflict with the study design); If possible, use experimental vehicles that can perform the respective driving scenario without restrictions.

Furthermore, we condense our learnings in a rough assessment of the discussed methods (Table 7.4). Particularly, we assess them in terms of their suitability for supporting the four key activities of the HCD process – i.e., understanding the context of use, specifying user requirements, producing design solutions, and evaluating designs – and regarding their level of immersion/realism, (external) validity, reliability, cost-efficiency, and options to combine them with other methods. Therefore, we use a simplified three-point rating scale ranging from low to high. Along with the recommendations listed in Table 7.4 and Table 7.5 may serve designers, researchers, and practitioners as a starting point for deciding on which method to use. However, we want to note that the assessment is rather fuzzy and the respective method's score and suitability may vary depending on influencing factors and corresponding variables (Table 7.3).

7.4 Reflections and Future Work

In the foreseeable future, AVs will find their way into our daily lives and continue to rise. So will the necessity to design suitable ways to interact with them. With our work, we want to contribute to developing desirable human-AV interactions by facilitating early prototyping activities. We argued that context-based interface prototyping should be used to define requirements, analyze acceptance and human factors challenges, and assess potential design solutions while considering their interrelations with contextual factors, particularly from early development phases. In the sense of HCAI (Section 2.2.2), we should aim for understandable, trustworthy, and secure human-AV interactions. We require a holistic perspective, also to counteract the discussed challenges. Furthermore, we conclude the need for (formalized/standardized) frameworks and toolkits for context-based prototyping to achieve valid and reliable assessments from early on. Based on our learnings, we outline what we consider relevant for future work regarding those two aspects and illustrate how future researchers and practitioners can build upon this dissertation's contributions.

7.4.1 Designing Reliable, Trustworthy, and Secure Human-AV Interactions

HCI designers strive to design optimal interactions between humans and machines. Apparently, no one has ever designed a "perfect" HMI that magically matches the needs of any possible user. This is simply because *it depends*. So is the case with human-AV interactions. Interactions between Humans and AVs are defined by a multitude of contextual and situational facets as well as inter- and intrapersonal characteristics (Figure 7.1). In line with the concepts of HCD and HCAI (Section 2.2.2), future work needs to (better) understand the interrelations of these factors and their consequent effects on (shared) AV/AMoD passengers' experiences more thoroughly and in a holistic manner. Based on that knowledge (and, of course, many testings and iterations), we may finally create HMIs for internal (and external) communication that provide users with just the right amount of information and functionality – and eventually, overcome the discussed human factors and acceptance challenges. We are not yet able to tell exactly how this should be done. However, based on our learnings, we want to point out some potential directions for future research on the way to designing desirable experiences by adopting a holistic perspective. Future work may build on those suggestions, refine the proposed approaches, and, finally, create reliable, trustworthy, and secure human-AV interactions.

Understandability and Trustworthiness by Design

From an HCAI/XAI perspective, intelligent systems should enable their users to understand them [204]. This may include facilitating humans to understand the AI-based systems' decisions – which are often still 'black boxes'. We saw that providing users with transparent feedback on AVs' reasoning can increase passengers' understanding (Chapter 4). While we implemented a rather basic visualization approach, future work may investigate more sophisticated and better integrated visualizations. But then again, we saw that providing (too much) system feedback can also have adverse effects and, e.g., distract or annoy users when it is too engaging. Consequently, future work must find the right balance between information provision, functionality, and users' context-dependent needs. Generally, future work should investigate interactions from a holistic perspective and create concepts that inherently provide users with understanding and (calibrated) trust toward the system by design.

Multimodal Interaction, Accessibility, and Configurability

The bandwidth of potential AV and AMoD users is vast. Consequently, this is also true for various individual needs, goals, skills, and preferences. Beyond safety- and security-critical events, AVs should provide users with multiple ways and modalities to interact with the autonomous system and facilitate user freedom. This could include 'classic' visual HMIs, but also (speech- and text-based) conversational UIs and auditory and tactile feedback. To ensure the accessibility of AVs and AMoD systems, future work should investigate appropriate HMIs for people with disabilities. Generally, providing users with multiple (redundant) ways to interact facilitates meeting accessibility requirements and promotes user freedom. In fact, future work may also consider adopting an ability-based approach [180] and provide, e.g., deaf and hard-hearing individuals with (avatar-based) sign language communication [180]. Building on our findings on mobile and in-vehicle HMIs (Chapters 5 and 6), future work may investigate the interplay of several modalities and touchpoints and the resulting 'hybrid' HMIs from a holistic perspective. Thereby, future work may also investigate passengers' demand to configure in-vehicle HMIs, e.g., in terms of system feedback (Chapter 4), and discover to what extent this should be possible and in which situations this may be of value in terms of passengers' acceptance and UX.

Balancing Security and Privacy Needs

In Chapter 6, we investigated the effects of providing shared AV users with information on their co-passengers. We saw that disclosing personal information on fellow passengers can decrease users' perceived risk and positively affect their UX, particularly during night rides. At the same time, participants raised privacy concerns associated with the provision of their own personal data to others. Future research should further investigate this matter and find the right balance between the antagonists security and privacy. Therefore, future work may investigate additional measures to increase (perceived) security and feeling of control of passengers and, e.g., consider (human) support/contact options before, during, and after the ride or concepts like artificial or human travel companions or buddies (e.g., [226]).

Long-term Effects and Summative Evaluations

Empirical research often relies on formative evaluations with 'single-use' tests. I.e., participants experience a product, system, or service in the respective user study for the first and only time. So is the case for our conducted studies (Chapters 3, 4, 5, and 6). This approach is quite common in HCI research and quite efficient, e.g., for evaluating early concepts and discovering usability issues. However, some effects from interacting with a product, system, or service become only identifiable after long-term monitoring [125]. Future work may follow the recommendation of ISO 9241-210 [125] and related work in the AV domain (e.g., Nordhoff et al. [181]) and investigate the designed (holistic) concepts for human-AV interaction also with long-term summative studies. For instance, Richardson [202] conducted a longitudinal [27] driving simulator study to investigate the long-term effects of an SAE level 3 system. Among other things, they found a significant increase in users' perceived usefulness over time. Conducting such studies for AVs may help to optimize acceptance and UX. However, they might also require further investigation, refinement, and possible standardization of methodological approaches and prototyping methods.

7.4.2 Toward a Decision Framework for Context-Based Prototyping of Human-AV Interactions

Our answers to RQ2 provide the foundation for a decision framework for context-based prototyping of human-AV interactions. They contain an overview of general influencing factors and variables (Table 7.3, Figure 7.1), a rough assessment of the discussed prototyping methods' suitability and strengths (Table 7.4) and practical recommendations for their application (Table 7.5). Future work may build on these contributions, validating the proposed foundations whilst also formalizing and extending them to a comprehensive decision and support framework for context-based prototyping of human-AV interactions. This may also include an extension of the prototyping recommendations and the creation of toolkits.

Validation of Framework Fundamentals

With Table 7.3 and Figure 7.1, we provide a collection of influencing factors and variables that can and – from our point of view – also should be considered in AV HMI design activities as early as possible. We also discussed their relationship and dependencies with Equation (7.1). Currently, though, this is a rather theoretical proposal. Future work should validate these fundamentals. Furthermore, each factor's practical relevance, matter, and 'weight' should be estimated. Practical implications might depend on the current stage of development, planned HCD activities, and respective use cases and scenarios of interest. Considering this, we derived a rough assessment of the methods discussed in this thesis based on our experiences with prototyping of in-vehicle HMIs (Table 7.4). While our assessment can serve as an initial basis to reason about the methods and their combinations and discuss their suitability for certain activities, it should be re-assessed and extended on a broader and more objective basis. For instance, systematic reviews/meta-analyses of user studies from related work featuring context-based prototyping may be a suitable approach.

Extension and Formalization of a Prototyping Decision Framework

Based on the proposed validation and assessment of its fundamentals, the framework might be formalized and provide (step-wise) guidance, decision support, and practical prototyping recommendations for researchers, designers, and practitioners. Future work may consider formalizing the framework as a step-by-step wizard to find the most suitable method(s) depending on relevant influencing factors and scenarios. Furthermore, future work may extend it by further investigating (and/or extending) influencing factors and their interrelations (Figure 7.1). Besides the physical and social context focused on in this thesis, this may also include a deeper exploration of cultural and situational aspects. A comprehensive formalized framework could support reliable and valid investigations and comparisons of HMIs, taking into account the various contextual factors that may be relevant.

Prototyping Recommendations and Toolkits

The prototyping recommendations we derived based on our learnings (Table 7.5) should be refined and validated – e.g., by applying them in empirical user studies. Furthermore, future work may extend the recommendations and the method collection with other (novel, combinations, or more granular contemplations of) suitable context-based prototyping methods. This might incorporate a comparative examination of the respective benefits and downsides of the methods for specific AV scenarios in a standardized manner. Lastly, the framework could be further extended to provide toolkits for applying specific methods and procedures. For instance, regarding video-based AV simulation (Chapter 3), such toolkits might provide essential work files and assets, including video footage and sound snippets, to create context-based AV (interface) prototypes. Furthermore, the toolkits might provide ways for collaboration among researchers, e.g., by exchanging footage and assets with peers.

7.4.3 Closing Remarks

AVs are right on our doorstep. Although not all technological challenges have been solved yet, AVs will likely become a major form of transportation in the not-so-far future. However, as we have seen, driverless vehicles face severe acceptance challenges among the population. These range from general skepticism or concerns regarding comfort and privacy to severe trust issues, safety and security concerns, and fear of AI-powered systems. This thesis investigated the design of HMIs suitable for counteracting these challenges. To address them early, we particularly argued for context-based prototyping of prospective human-AV interactions and derived recommendations for future work. We hope that our work contributes its share to overcome the hurdles and that – if humanity's dream of driverless and accident-free travel (Figure 1.1) finally comes true on a large scale – the technology can realize its full potential. We live in exciting times and look forward to what the autonomous future of mobility holds.

Vielleicht, wer woaß des gwieß.
Es konn sei, dass moang d’Weid untergeht,
oda dass ois so bleibt wia’s is.

Pam Pam Ida, ‘Vielleicht’

Acronyms

aHMI automation HMI.

AI artificial intelligence.

AMoD autonomous mobility-on-demand.

ANOVA analysis of variance.

AR augmented reality.

ATI affinity for technology interaction.

AV autonomous vehicle.

CAVE Cave Automatic Virtual Environment.

CCTV closed circuit television.

CGI computer-generated imagery.

CoP change of plans.

CUI conversational user interface.

dHMI dynamic HMI.

eHMI external HMI.

GUI graphical user interface.

HCAI human-centered artificial intelligence.

HCD human-centered design.

HCI human-computer interaction.

HMI human-machine interface.

iHMI infotainment HMI.

IPQ Igroup Presence Questionnaire.

ISO International Organization for Standardization.

MR mixed reality.

MSAQ Motion Sickness Assessment Questionnaire.

PE perceived ease of use.

PU perceived usefulness.

RM-ANOVA repeated measures analysis of variance.

RQ research question.

SAE Society of Automotive Engineers.

SAMoD shared autonomous mobility-on-demand.

SAV shared autonomous vehicle.

TAM Technology Acceptance Model.

TiA trust in automation.

TICS transport information and control systems.

TV television.

UEQ User Experience Questionnaire.

UEQ-S User Experience Questionnaire, short version.

UI user interface.

UX user experience.

vHMI vehicle HMI.

VR virtual reality.

WoOz Wizard-of-Oz.

XAI explainable artificial intelligence.

List of Figures

List of Tables

Bibliography

[1] Tousif Ahmed, Apu Kapadia, Venkatesh Potluri, and Manohar Swaminathan. 2018. Up to a Limit? Privacy Concerns of Bystanders and Their Willingness to Share Additional Information with Visually Impaired Users of Assistive Technologies. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 89 (sep 2018), 27 pages. `https://doi.org/10.1145/3264899`

[2] Tony Ahn, Seewon Ryu, and Ingoo Han. 2007. The impact of Web quality and playfulness on user acceptance of online retailing. *Information & Management* 44, 3 (2007), 263–275. `https://doi.org/10.1016/j.im.2006.12.008`

[3] Deike Albers, Jonas Radlmayr, Alexandra Loew, Sebastian Hergeth, Frederik Naujoks, Andreas Keinath, and Klaus Bengler. 2020. Usability Evaluation – Advances in Experimental Design in the Context of Automated Driving Human-Machine Interfaces. *Information* 11, 5 (2020), 15 pages. `https://doi.org/10.3390/INFO11050240`

[4] Paul D. Allison. 2009. Missing Data. In *SAGE Handbook of Quantitative Methods in Psychology*, Roger E. Millsap and Alberto Maydeu-Olivares (Eds.). Sage Publications Inc., Thousand Oaks, CA, 72–89. `https://doi.org/10.4135/9780857020994`

[5] Mustafa Almallah, Qinaat Hussain, Nora Reinolsmann, and Wael K.M. Alhajyaseen. 2021. Driving simulation sickness and the sense of presence: Correlation and contributing factors. *Transportation Research Part F: Traffic Psychology and Behaviour* 78 (2021), 180–193. `https://doi.org/10.1016/j.trf.2021.02.005`

[6] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. `https://doi.org/10.1145/3290605.3300233`

[7] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. 2001. Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications* 21, 6 (2001), 34–47. `https://doi.org/10.1109/38.963459`

[8] Baidu. 2020. Building a self-driving car that people can trust. Retrieved 2023-08-02 from `https://www.technologyreview.com/2020/12/16/1014672/building-a-self-driving-car-that-people-can-trust/`

[9] William Sims Bainbridge. 2004. Berkshire Encyclopedia of Human-Computer Interaction. In *Berkshire Encyclopedia of Human-Computer Interaction*, William Sims Bainbridge (Ed.). Vol. 1. Berkshire Publishing Group LLC, Massachusetts, xxxiii–xxxix. `https://www.berkshirepublishing.com/title/berkshire-encyclopedia-of-human-computer-interaction/`

[10] Saskia Bakker and Karin Niemantsverdriet. 2016. The interaction-attention continuum : considering various levels of human attention in interaction design. *International Journal of Design* 10, 2 (2016), 1–14.

[11] Sonia Baltodano, Srinath Sibi, Nikolas Martelaro, Nikhil Gowda, and Wendy Ju. 2015. The RRADS Platform: A Real Road Autonomous Driving Simulator. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Nottingham, United Kingdom) *(AutomotiveUI '15)*. Association for Computing Machinery, New York, NY, USA, 281–288. `https://doi.org/10.1145/2799250.2799288`

[12] Richard Bartle. 2003. *Designing Virtual Worlds.* New Riders, San Francisco, CA, USA. `https://dl.acm.org/doi/10.5555/1196681`

[13] Arthur Barz, Jan Conrad, and Dieter Wallach. 2020. Advantages of Using Runtime Procedural Generation of Virtual Environments Based on Real World Data for Conducting Empirical Automotive Research. In *Proceedings of the 22nd International Conference on Human-Computer Interaction* (Copenhagen, Denmark) *(HCII '20)*. Springer International Publishing, Cham, Switzerland, 14–23. `https://doi.org/10.1007/978-3-030-59987-4_2`

[14] Dirk Bäumer, Walter R. Bischofberger, Horst Lichter, and Heinz Züllighoven. 1996. User Interface Prototyping – Concepts, Tools, and Experience. In *Proceedings of the 18th International Conference on Software Engineering* (Berlin, Germany) *(ICSE '96, March)*. Institute of Electrical and Electronics Engineers, New York, NY, USA, 532–541. `https://doi.org/10.1109/ICSE.1996.493447`

[15] Sven Beiker, Frederik Hansson, Anders Suneson, and Michael Uhl. 2016. How the convergence of automotive and tech will create a new ecosystem. Retrieved 2022-10-11 from `https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/how-the-convergence-of-automotive-and-tech-will-create-a-new-ecosystem`

[16] Victoria Bellotti and Keith Edwards. 2001. Intelligibility and Accountability: Human Considerations in Context-Aware Systems. *Human–Computer Interaction* 16, 2-4 (2001), 193–212. `https://doi.org/10.1207/S15327051HCI16234_05`

[17] Klaus Bengler. 2015. Messmethoden: Notwendigkeit von Versuchen. In *Automobilergonomie*, Heiner Bubb, Klaus Bengler, Rainer E. Grünen, and Mark Vollrath (Eds.). Springer Fachmedien, Wiesbaden, Germany, Chapter 11, 618–621. `https://doi.org/10.1007/978-3-8348-2297-0_11`

[18] Klaus Bengler, Klaus Dietmayer, Berthold Färber, Markus Maurer, Christoph Stiller, and Hermann Winner. 2014. Three Decades of Driver Assistance Systems – Review and Future Perspectives. *IEEE Intelligent Transportation Systems Magazine* 6, 4 (2014), 6–22. `https://doi.org/10.1109/MITS.2014.2336271`

[19] Klaus Bengler, Kamil Omozik, and Andrea Isabell Müller. 2019. The Renaissance of Wizard of Oz (WoOz) - Using the WoOz methodology to prototype automated vehicles. In *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2019 Annual Conference* (Nantes, France) *(HFES '19)*. Human Factors and Ergonomics Society, Europe Chapter, Liverpool, England UK, 63–72. `https://www.researchgate.net/publication/346659448`

[20] Klaus Bengler, Michael Rettenmaier, Nicole Fritz, and Alexander Feierle. 2020. From HMI to HMIs: Towards an HMI framework for automated driving. *Information* 11, 61 (2020), 17. `https://doi.org/10.3390/info11020061`

[21] Ruha Benjamin. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity / Wiley & Sons, Hoboken, NJ, USA. `https://www.wiley.com/en-us/Race+After+Technology:+Abolitionist+Tools+for+the+New+Jim+Code-p-9781509526437`

[22] Melanie Berger, Aditya Dandekar, Regina Bernhaupt, and Bastian Pfleging. 2021. An AR-Enabled Interactive Car Door to Extend In-Car Infotainment Systems for Rear Seat Passengers. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 404, 6 pages. `https://doi.org/10.1145/3411763.3451589`

[23] Niels Ole Bernsen, Hans Dybkjær, and Laila Dybkjær. 1994. Wizard of Oz Prototyping: When and How? *CCI Working Papers in Cognitive Science and HCI* (1994), 1–13. `http://spokendialogue.dk/Publications/1994i/WPCS-94-1-10.2.94.pdf`

[24] Hannah Biermann, Ralf Philipsen, Teresa Brell, and Martina Ziefle. 2020. Shut Up and Drive? User Requirements for Communication Services in Autonomous Driving. In *Proceedings of the 22nd International Conference on Human-Computer Interaction* (Copenhagen, Denmark) *(HCII '20)*. Springer, Cham, Switzerland, 3–14. `https://doi.org/10.1007/978-3-030-50523-3_1`

[25] Jan Blom, Divya Viswanathan, Mirjana Spasojevic, Janet Go, Karthik Acharya, and Robert Ahonius. 2010. Fear and the City: Role of Mobile Services in Harnessing Safety and Security in Urban Use Contexts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1841–1850. `https://doi.org/10.1145/1753326.1753602`

[26] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* 2004.10934 (apr 2020), 17 pages. `https://doi.org/10.48550/arXiv.2004.10934`

[27] Jürgen Bortz and Nicola Döring. 2006. *Forschungsmethoden und Evaluation* (4 ed.). Springer, Berlin and Heidelberg, Germany. `https://doi.org/10.1007/978-3-540-33306-7`

[28] Richard E. Boyatzis. 1998. *Transforming Qualitative Information: Thematic Analysis and Code Development*. Sage Publications, Thousand Oaks, CA, USA.

[29] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. `https://doi.org/10.1191/1478088706qp063oa`

[30] Teresa Brell, Hannah Biermann, Ralf Philipsen, and Martina Ziefle. 2019. Conditional Privacy: Users' Perception of Data Privacy in Autonomous Driving. In *Proceedings of the 5th International Conference on Vehicle Technology and Intelligent Transport Systems* (Heraklion, Crete, Greece) *(VEHITS '19)*. Science and Technology Publications, Setúbal, Portugal., 352–359. `https://doi.org/10.5220/0007693803520359`

[31] Teresa Brell, Ralf Philipsen, and Martina Ziefle. 2019. Suspicious minds? – users ' perceptions of autonomous and connected driving. *Theoretical Issues in Ergonomoics Science* 20, 3 (2019), 301–331. `https://doi.org/10.1080/1463922X.2018.1485985`

[32] Ralph Bruder and Muriel Didier. 2015. Gestaltung von Mensch-Maschine-Schnittstellen. In *Handbuch Fahrerassistenzsysteme* (3 ed.), H. Winner, S. Hakuli, F. Lotz, and C. Singer (Eds.). Springer Vieweg, Wiesbaden, Germany, Chapter 34, 633 – 646. `https://doi.org/10.1007/978-3-658-05734-3`

[33] Heiner Bubb. 2015. Methoden der ergonomischen Fahrzeugentwicklung. In *Automobilergonomie*, Heiner Bubb, Klaus Bengler, Rainer E. Grünen, and Mark Vollrath (Eds.). Springer Fachmedien, Wiesbaden, Germany, Chapter 10, 583 – 617. `https://doi.org/10.1007/978-3-8348-2297-0_10`

[34] Heiner Bubb, Klaus Bengler, Jurek Breuninger, Christian Gold, and Magnus Helmbrecht. 2015. Systemergonomie des Fahrzeugs. In *Automobilergonomie*, Heiner Bubb, Klaus Bengler, Rainer E. Grünen, and Mark Vollrath (Eds.). Springer Fachmedien, Wiesbaden, Germany, Chapter 6, 259–344. `https://doi.org/10.1007/978-3-8348-2297-0_6`

[35] Marion Buchenau and Jane Fulton Suri. 2000. Experience Prototyping. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (New York, NY, USA) *(DIS '00)*. Association for Computing Machinery, New York, NY, USA, 424–433. `https://doi.org/10.1145/347642.347802`

[36] Raluca Budiu. 2018. Between-Subjects vs. Within-Subjects Study Design. Retrieved 2022-09-01 from `https://www.nngroup.com/articles/between-within-subjects/`

[37] Raluca Budiu. 2018. The User Experience of Chatbots. Retrieved 2020-10-02 from `https://www.nngroup.com/articles/chatbots/`

[38] Bill Buxton. 2007. Multi-Touch Systems that I Have Known and Loved. Retrieved 2022-10-17 from `www.billbuxton.com/multitouchOverview.html`

[39] Cambridge Dictionary. 2022. Meaning of context in English. Retrieved 2022-10-20 from `https://dictionary.cambridge.org/us/dictionary/english/context`

[40] Bradley Camburn, Vimal Viswanathan, Julie Linsey, David Anderson, Daniel Jensen, Richard Crawford, Kevin Otto, and Kristin Wood. 2017. Design prototyping methods: state of the art in strategies, techniques, and guidelines. *Design Science* 3, 13 (2017), 1–33. `https://doi.org/10.1017/dsj.2017.10`

[41] Ching-Fu Chen. 2019. Factors affecting the decision to use autonomous shuttle services: Evidence from a scooter-dominant urban context. *Transportation Research Part F: Traffic Psychology and Behaviour* 67 (2019), 195–204. `https://doi.org/10.1016/j.trf.2019.10.016`

[42] Vincent Cho, T.C. Edwin Cheng, and W.M. Jennifer Lai. 2009. The role of perceived user-interface design in continued usage intention of self-paced e-learning tools. *Computers & Education* 53, 2 (2009), 216–227. `https://doi.org/10.1016/j.compedu.2009.01.014`

[43] Jong Kyu Choi and Yong Gu Ji. 2015. Investigating the Importance of Trust on Adopting an Autonomous Vehicle. *International Journal of Human-Computer Interaction* 31, 10 (2015), 692–702. `https://doi.org/10.1080/10447318.2015.1070549`

[44] Zhuang Jie Chong, Baoxing Qin, Tirthankar Bandyopadhyay, Tichakorn Wongpiromsarn, Brice Rebsamen, P. Dai, E. S. Rankin, and Marcelo H. Ang. 2013. Autonomy for Mobility on Demand. In *Intelligent Autonomous Systems 12*, Sukhan Lee, Hyungsuck Cho, Kwang-Joon Yoon, and Jangmyung Lee (Eds.). Springer, Berlin and Heidelberg, Germany, 671–682. `https://doi.org/10.1007/978-3-642-33926-4_64`

[45] William Clayton, Daniela Paddeu, Graham Parkhurst, and John Parkin. 2020. Autonomous Vehicles: Who will use them, and will they share? *Transportation Planning and Technology* 43, 4 (2020), 343–364. `https://doi.org/10.1080/03081060.2020.1747200`

[46] Jacob Cohen. 1992. A Power Primer. *Psychological Bulletin* 112, 1 (1992), 155–159. `https://doi.org/10.1037/0033-2909.112.1.155`

[47] Mark Colley, Benjamin Eder, Jan Ole Rixen, and Enrico Rukzio. 2021. Effects of Semantic Segmentation Visualization on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 155, 11 pages. `https://doi.org/10.1145/3411764.3445351`

[48] Mark Colley, Max Rädler, Jonas Glimmann, and Enrico Rukzio. 2022. Effects of Scene Detection, Scene Prediction, and Maneuver Planning Visualizations on Trust, Situation Awareness, and Cognitive Load in Highly Automated Vehicles. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 49 (jul 2022), 21 pages. `https://doi.org/10.1145/3534609`

[49] Condens Insights GmbH. 2022. Supercharge your UX research analysis with condens. Retrieved 2022-08-24 from `https://condens.io/`

[50] Andy Crabtree. 2003. *Designing Collaborative Systems – A Practical Guide to Ethnography*. Springer, London, England UK. `https://doi.org/10.1007/b97516`

[51] Henriette Cramer and Juho Kim. 2019. Confronting the Tensions where UX Meets AI. *Interactions* 26, 6 (2019), 69–71. `https://doi.org/10.1145/3364625`

[52] John W. Creswell. 2014. *A Concise Introduction to Mixed Methods Research*. Sage Publications, Thousand Oaks, CA, USA. `https://us.sagepub.com/hi/nam/a-concise-introduction-to-mixed-methods-research/book266037`

[53] Cruise. 2021. Cruise Passenger Safety Plan. Retrieved 2023-08-02 from `https://www.cpuc.ca.gov/-/media/cpuc-website/divisions/consumer-protection-and-enforcement-division/documents/tlab/av-programs/phase-i-av-deployment-program-al-status/2021-11-05-advice-letter---cruise-application-for-driverless-deployment-permit.pdf`

[54] Cruise. 2023. April 2023 software release. Retrieved 2023-08-02 from `https://getcruise.com/news/blog/2023/april-2023-software-release/`

[55] Carolina Cruz-Neira, Daniel J. Sandin, and Thomas A. DeFanti. 1993. Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques* (Anaheim, CA, USA) *(SIGGRAPH '93)*. Association for Computing Machinery, New York, NY, USA, 135–142. `https://doi.org/10.1145/166117.166134`

[56] Graham Currie. 2018. Lies, Damned Lies, AVs, Shared Mobility, and Urban Transit Futures. *Journal of Public Transportation* 21, 1 (2018), 19–30. `https://doi.org/10.5038/2375-0901.21.1.3`

[57] Graham Currie, Alexa Delbosc, and Sarah Mahmoud. 2013. Factors Influencing Young Peoples' Perceptions of Personal Safety on Public Transport. *Journal of Public Transportation* 16, 1 (2013), 1–19. `https://doi.org/10.5038/2375-0901.16.1.1`

[58] Yngve Dahl, Ole Andreas Alsos, and Dag Svanæs. 2009. Evaluating Mobile Usability: The Role of Fidelity in Full-Scale Laboratory Simulations with Mobile ICT for Hospitals. In *Proceedings of the 13th International Conference on Human-Computer Interaction* (San Diego, CA, USA) *(HCII '09)*. Springer, Berlin and Heidelberg, Germany, 232–241. `https://doi.org/10.1007/978-3-642-02574-7_26`

[59] Peter Dalsgaard. 2010. Research in and through Design: An Interaction Design Research Approach. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction* (Brisbane, Australia) *(OZCHI '10)*. Association for Computing Machinery, New York, NY, USA, 200–203. `https://doi.org/10.1145/1952222.1952265`

[60] Fred D. Davis. 1986. *A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results*. Dissertation. Massachusetts Institute of Technology. `http://hdl.handle.net/1721.1/15192`

[61] Joost C. F. de Winter, P. M. van Leeuwen, and Riender Happee. 2012. Advantages and Disadvantages of Driving Simulators: A Discussion. In *Proceedings of Measuring Behavior 2012* (Utrecht, Netherlands). Noldus Information Technology, Wageningen, Netherlands, 47–50. `https://research.tudelft.nl/en/publications/advantages-and-disadvantages-of-driving-simulators-a-discuossion`

[62] Marissa Janine Dekker. 2017. *Riding a self-driving bus to work: Investigating how travellers perceive ADS-DVs on the last mile*. Master's Thesis. Delf University of Technology, Delft, Netherlands. `http://resolver.tudelft.nl/uuid:c9773355-d4a3-4063-9799-11343c625c01`

[63] Henrik Detjen, Sarah Faltaous, Bastian Pfleging, Stefan Geisler, and Stefan Schneegass. 2021. How to Increase Automated Vehicles' Acceptance through In-Vehicle Interaction Design: A Review. *International Journal of Human-Computer Interaction* 37, 4 (2021), 308–330. `https://doi.org/10.1080/10447318.2020.1860517`

[64] Henrik Detjen, Bastian Pfleging, and Stefan Schneegass. 2020. A Wizard of Oz Field Study to Understand Non-Driving-Related Activities, Trust, and Acceptance of Automated Vehicles. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Virtual Event, DC, USA) *(AutomotiveUI '20)*. Association for Computing Machinery, New York, NY, USA, 19–29. `https://doi.org/10.1145/3409120.3410662`

[65] Anind K. Dey and Gregory D. Abowd. 1999. *Towards a Better Understanding of Context and Context-Awareness*. GVU Center Technical Reports. Georgia Tech Library, Atlanta, GA, USA. `http://hdl.handle.net/1853/3389`

[66] Debargha Dey and Jacques Terken. 2017. Pedestrian Interaction with Vehicles: Roles of Explicit and Implicit Communication. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Oldenburg, Germany) *(AutomotiveUI '17)*. Association for Computing Machinery, New York, NY, USA, 109–113. `https://doi.org/10.1145/3122986.3123009`

[67] Klaus Dietmayer. 2016. Predicting of Machine Perception for Automated Driving. In *Autonomous Driving: Technical, Legal and Social Aspects*, Markus Maurer, J. Christian Gerdes, Barbara Lenz, and Hermann Winner (Eds.). Springer, Berlin and Heidelberg, Germany, 407–424. `https://doi.org/10.1007/978-3-662-48847-8_20`

[68] A. Dillon. 2001. Human Acceptance of Information Technology. In *International Encyclopedia of Ergonomics and Human Factors* (1 ed.), Waldemar Karwowski (Ed.). Taylor & Francis, London, England UK and New York, NY, USA, 673–675. `https://doi.org/10.1201/9781482298536`

[69] Verena Distler, Carine Lallemand, and Thierry Bellet. 2018. Acceptability and Acceptance of Autonomous Mobility on Demand: The Impact of an Immersive Experience. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. `https://doi.org/10.1145/3173574.3174186`

[70] Mike Dodge. 2019. A Scientific Method for Impactful UX Prototyping – And why 'rapid prototyping' isn't enough to critically impact product design. Retrieved 2022-10-17 from `https://modus.medium.com/a-scientific-method-for-impactful-ux-prototyping-9ab6c7a1dcde`

[71] Scooter Doll. 2022. Uber deploying Motional IONIQ 5 driverless robotaxis to millions of riders over the next decade. Retrieved 2022-10-17 from `https://electrek.co/2022/10/06/uber-deploying-motional-ioniq-5-robotaxis-to-millions-of-riders/`

[72] Pedro M. D'Orey, Amin Hosseini, Jose Azevedo, Frank Diermeyer, Michel Ferreira, and Markus Lienkamp. 2016. Hail-a-Drone: Enabling teleoperated taxi fleets. In *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '16)* (Gothenburg, Sweden). Institute of Electrical and Electronics Engineers, New York, NY, USA, 774–781. `https://doi.org/10.1109/IVS.2016.7535475`

[73] Jennifer Dungs, Florian Herrmann, Daniel Duwe, Alexander Schmidt, Sebastian Stegmüller, Ralf Gaydoul, Peter Lukas Peters, and Maik Sohl. 2016. The Value of Time – Nutzerbezogene Service-Potenziale durch autonomes Fahren. Retrieved 2019-12-12 from `https://blog.iao.fraunhofer.de/images/blog/studie-value_of_time.pdf`

[74] Grace Eden. 2018. Transforming cars into computers: Interdisciplinary opportunities for HCI. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference* (Swindon, England UK) *(HCI '18)*. BCS Learning and Development, Belfast, Northern Ireland UK, 3 pages. `https://doi.org/10.14236/ewic/HCI2018.73`

[75] e.GO Mobile AG. 2020. e.GO Mover. Retrieved 2020-01-14 from `https://www.e-go-mobile.com/en/models/e.go-mover/`

[76] Mica R. Endsley. 1995. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors* 37, 1 (1995), 32–64. `https://doi.org/10.1518/001872095779049543`

[77] Mica R. Endsley and Esin O. Kiris. 1995. The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors* 37, 2 (1995), 381–394. `https://doi.org/10.1518/001872095779064555`

[78] European Commission. 2015. Vision Zero. Retrieved 2023-04-25 from `https://road-safety.transport.ec.europa.eu/index_en`

[79] Gary W. Evans and Richard Wener. 2007. Crowding and personal space invasion on the train: Please don't make me sit in the middle. *Journal of Environmental Psychology* 27, 1 (2007), 90–94. `https://doi.org/10.1016/j.jenvp.2006.10.002`

[80] Zhihan Fang, Boyang Fu, Zhou Qin, Fan Zhang, and Desheng Zhang. 2020. PrivateBus: Privacy Identification and Protection in Large-Scale Bus WiFi Systems. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 9 (mar 2020), 23 pages. `https://doi.org/10.1145/3380990`

[81] Federal Ministry for Digital and Transport. 2021. Germany will be the world leader in autonomous driving. Retrieved 2022-10-17 from `https://bmdv.bund.de/SharedDocs/EN/Articles/DG/act-on-autonomous-driving.html`

[82] Alexander Feierle, Simon Danner, Sarah Steininger, and Klaus Bengler. 2020. Information Needs and Visual Attention during Urban, Highly Automated Driving—An Investigation of Potential Influencing Factors. *Information* 11, 2 (2020), 1–17. `https://doi.org/10.3390/info11020062`

[83] Frank Flemisch, Johann Kelsch, Christian Löper, Anna Schieben, and Julian Schindler. 2008. Automation spectrum, inner / outer compatibility and other potentially useful human factors concepts for assistance and automation. In *Human Factors for assistance and automation*, D. de Waard, F. O. Flemisch, B. Lorenz, H. Oberheid, and K. A. Brookhuis (Eds.). Shaker Publishing, Maastricht, Netherlands, 1–16. `https://elib.dlr.de/57625/`

[84] Fabian Flohr, Madalin Dumitru-Guzu, Julian F. P. Kooij, and Dariu M. Gavrila. 2015. A Probabilistic Framework for Joint Pedestrian Head and Body Orientation Estimation. *IEEE Transactions on Intelligent Transportation Systems* 16, 4 (2015), 1872–1882. `https://doi.org/10.1109/TITS.2014.2379441`

[85] Lukas A. Flohr, Dominik Janetzko, Dieter P. Wallach, Sebastian C. Scholz, and Antonio Krüger. 2020. Context-Based Interface Prototyping and Evaluation for (Shared) Autonomous Vehicles Using a Lightweight Immersive Video-Based Simulator. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1379–1390. `https://doi.org/10.1145/3357236.3395468`

[86] Lukas A. Flohr, Sofie Kalinke, Antonio Krüger, and Dieter P. Wallach. 2021. Chat or Tap? – Comparing Chatbots with 'Classic' Graphical User Interfaces for Mobile Interaction with Autonomous Mobility-on-Demand Systems. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 21, 13 pages. `https://doi.org/10.1145/3447526.3472036`

[87] Lukas A. Flohr, Martina Schuß, Dieter Wallach, Antonio Krüger, and Andreas Riener. 2023. Designing for Passengers' Information Needs on Fellow Travelers: A Comparison of Day and Night Rides in Shared Automated Vehicles. *arXiv* 2308.02616 (aug 2023), 22 pages. `https://doi.org/10.48550/arXiv.2308.02616`

[88] Lukas A. Flohr, Joseph S. Valiyaveettil, Antonio Krüger, and Dieter Wallach. 2023. Prototyping Autonomous Vehicle Windshields with AR and Real-Time Object Detection Visualization: An On-Road Wizard-of-Oz Study. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. `https://doi.org/10.1145/3563657.3596051`

[89] Lukas A. Flohr and Dieter P. Wallach. 2023. The Value of Context-Based Interface Prototyping for the Autonomous Vehicle Domain: A Method Overview. *Multimodal Technologies and Interaction* 7, 4 (2023), 1–17. `https://doi.org/10.3390/mti7010004`

[90] Asbjørn Følstad and Petter Bae Brandtzæg. 2017. Chatbots and the New World of HCI. *Interactions* 24, 4 (2017), 38–42. `https://doi.org/10.1145/3085558`

[91] Asbjørn Følstad and Petter Bae Brandtzaeg. 2020. Users' experiences with chatbots: findings from a questionnaire study. *Quality and User Experience* 5, 3 (2020), 1–14. `https://doi.org/10.1007/s41233-020-00033-2`

[92] Asbjørn Følstad and Marita Skjuve. 2018. Business and pleasure? Relational interaction in conversational UX. Retrieved 2023-05-28 from `https://sintef.brage.unit.no/sintef-xmlui/handle/11250/2500859`

[93] Eva Fraedrich, Rita Cyganski, Ingo Wolf, and Barbara Lenz. 2016. *User Perspectives on Autonomous Driving – A Use-Case-Driven Study in Germany*. Technical Report. Geographisches Institut, Humboldt-Universität zu Berlin, Berlin, Germany. `https://doi.org/10.13140/RG.2.1.2539.4969`

[94] Thomas Franke, Christiane Attig, and Daniel Wessel. 2019. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human-Computer Interaction* 35, 6 (2019), 456–467. `https://doi.org/10.1080/10447318.2018.1456150`

[95] Peter Fröhlich, Raimund Schatz, Markus Buchta, Johann Schrammel, Stefan Suette, and Manfred Tscheligi. 2019. "What's the Robo-Driver up to?" Requirements for Screen-based Awareness and Intent Communication in Autonomous Buses. *i-com* 18, 2 (2019), 151–165. `https://doi.org/10.1515/icom-2018-0032`

[96] Tanja Fuest. 2021. *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*. Dissertation. Technical University of Munich, Munich, Germany. `https://mediatum.ub.tum.de/?id=1609297`

[97] Generated Media. 2019. Generated Photos: Faces. Retrieved 2023-03-18 from `https://generated.photos/faces`

[98] Michael A. Gerber, Ronald Schroeter, and Julia Vehns. 2019. A Video-Based Automated Driving Simulator for Automotive UI Prototyping, UX and Behaviour Research. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 14–23. `https://doi.org/10.1145/3342197.3344533`

[99] Lars Gerhold (Ed.). 2020. *Sicherheitsempfinden, Sicherheitskommunikation und Sicherheitsmaßnahmen: Ergebnisse aus dem Forschungsverbund WiSima*. Forschungsforum Öffentliche Sicherheit, Berlin, Germany. `https://doi.org/10.17169/refubium-27261`

[100] German Research Center for Artificial Intelligence (DFKI). 2019. OpenDS: Open Source Driving Simulation. Retrieved 2019-09-18 from `https://opends.dfki.de/`

[101] Peter J. Gianaros, Eric R. Muth, J. Toby Mordkoff, Max E. Levine, and Robert M. Stern. 2001. A Questionnaire for the Assessment of the Multiple Dimensions of Motion Sickness. *Aviat. Space Environ. Med.* 72, 2 (2001), 115–119. `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2910410/`

[102] Ulrich Gnewuch, Stefan Morana, Marc T.P. Adam, and Alexander Maedche. 2018. Faster is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction. In *Proceedings of the 26th European Conference on Information Systems* (Portsmouth, England UK) *(ECIS '18)*. Association for Information Systems, Atlanta, Georgia, USA. `https://aisel.aisnet.org/ecis2018_rp/113`

[103] Google Developers. 2020. Conversation design | Dialogflow and legacy Actions SDK | Google Developers. Retrieved 2020-09-09 from `https://developers.google.com/assistant/df-asdk/design`

[104] Greenwich Automated Transport Environment. 2018. GATEway Project: Final Report – This is Just the Beginning. Retrieved 2023-04-15 from `https://trl.co.uk/Uploads/TRL/Documents/D1.3_GATEway-Project-Final-Report-brochure.pdf`

[105] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang Zhong Yang. 2019. XAI-Explainable artificial intelligence. *Science Robotics* 4, 37 (2019), 4 pages. `https://doi.org/10.1126/scirobotics.aay7120`

[106] Chana J. Haboucha, Robert Ishaq, and Yoram Shiftan. 2017. User preferences regarding autonomous vehicles. *Transportation Research Part C: Emerging Technologies* 78 (2017), 37–49. `https://doi.org/10.1016/j.trc.2017.01.010`

[107] Erika Hall. 2018. *Conversational Design*. A Book Apart, Sebastopol, CA. `https://www.oreilly.com/library/view/conversational-design/9781492017950/`

[108] Catherine Harvey and Neville A. Stanton. 2013. A Usability Evaluation Framework for In-Vehicle Information Systems. In *Usability Evaluation for In-Vehicle Systems*. CRC Press, Boca Raton, Chapter 4, 49–70. `https://doi.org/10.1201/b14644`

[109] Marc Hassenzahl, Michael Burmester, and Franz Koller. 2003. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch und Computer 2003. Interaktion in Bewegung*, J. Ziegler and G. Szwillus (Eds.). B.G. Teubner, Stuttgart, Leipzig, 187–196. `https://doi.org/10.1007/978-3-322-80058-9_19`

[110] Marc Hassenzahl, Sarah Diefenbach, and Anja Göritz. 2010. Needs, affect, and interactive products - Facets of user experience. *Interacting with Computers* 22, 5 (2010), 353–362. `https://doi.org/10.1016/j.intcom.2010.04.002`

[111] Renate Häuslschmid, Max von Bülow, Bastian Pfleging, and Andreas Butz. 2017. SupportingTrust in Autonomous Driving. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus) *(IUI '17)*. Association for Computing Machinery, New York, NY, USA, 319–329. `https://doi.org/10.1145/3025171.3025198`

[112] Andrew J. Hawkins. 2019. Waymo's self-driving cars are now available on Lyft's app in Phoenix - The Verge. Retrieved 2019-09-14 from `https://www.theverge.com/2019/5/7/18536003/waymo-lyft-self-driving-ride-hail-app-phoenix`

[113] Ann-Christin Hensch, Isabel Neumann, Matthias Beggiato, Josephine Halama, and Josef F. Krems. 2019. 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 (2019), 18–32. `https://doi.org/10.5507/TOTS.2019.012`

[114] Sampo Hietanen. 2014. 'Mobility as a Service' – the new transport model? *Eurotransport* 12, 2 (2014), 2–4. `https://silo.tips/download/sampo-hietanen-ceo-its-finland`

[115] Henning Hinderer, Jonas Stegmüller, Jannick Schmidt, Jessica Sommer, and Jennifer Lucke. 2018. Acceptance of Autonomous Vehicles in Suburban Public Transport. In *Proceedings of the 2018 IEEE International Conference on Engineering, Technology and Innovation* (Stuttgart, Germany) *(ICE/ITMC '18)*. Institute of Electrical and Electronics Engineers, New York, NY, USA, 8 pages. `https://doi.org/10.1109/ICE.2018.8436261`

[116] Philipp Hock, Johannes Kraus, Franziska Babel, Marcel Walch, Enrico Rukzio, and Martin Baumann. 2018. How to Design Valid Simulator Studies for Investigating User Experience in Automated Driving: Review and Hands-On Considerations. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Toronto, ON, Canada) *(AutomotiveUI '18)*. Association for Computing Machinery, New York, NY, USA, 105–117. `https://doi.org/10.1145/3239060.3239066`

[117] Marius Hoggenmüller. 2022. *Urban Robotic Interfaces: Designing for Encounters with Non-Humanoid Robots in Cities*. Dissertation. The University of Sydney, Sydney, Australia. `https://hdl.handle.net/2123/29327`

[118] Marius Hoggenmüller, Martin Tomitsch, Luke Hespanhol, Tram Thi Minh Tran, Stewart Worrall, and Eduardo Nebot. 2021. Context-Based Interface Prototyping: Understanding the Effect of Prototype Representation on User Feedback. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 370, 14 pages. `https://doi.org/10.1145/3411764.3445159`

[119] Eric Horvitz. 1999. Principles of Mixed-Initiative User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 159–166. `https://doi.org/10.1145/302979.303030`

[120] International Transport Forum (ITF). 2018. *Women's Safety and Security – A Public Transport Priority*. Technical Report. International Transport Forum, Paris, France. 1–45 pages. `https://www.itf-oecd.org/sites/default/files/docs/womens-safety-security_0.pdf`

[121] ISO 15005. 2017. Road vehicles - Ergonomic aspects of transportation and control systems - Dialogue management principles and compliance procedures (ISO 15005:2017); German version EN ISO 15005:2017. `https://doi.org/10.31030/2584364`

[122] ISO 9241-110. 2008. Ergonomics of human-system interaction – Part 110: Dialogue principles (ISO 9241-110:2006) English version of DIN EN ISO 9241-110:2008-09. `https://doi.org/10.31030/1464024`

[123] ISO 9241-171. 2008. Ergonomics of human-system interaction – Part 171: Guidance on software accessibility (ISO 9241-171:2008) English version of DIN EN ISO 9241-171:2008-10. `https://doi.org/10.31030/1426650`

[124] ISO 9241-210. 2011. Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems (ISO 9241-210:2010) English translation of DIN EN ISO 9241-210:2011-01. `https://doi.org/10.31030/1728173`

[125] ISO 9241-210. 2020. Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems (ISO 9241-210:2019) German version EN ISO 9241-210:2019. `https://doi.org/10.31030/3104744`

[126] Pascal Jansen, Mark Colley, and Enrico Rukzio. 2022. A Design Space for Human Sensor and Actuator Focused In-Vehicle Interaction Based on a Systematic Literature Review. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2, Article 56 (jul 2022), 51 pages. `https://doi.org/10.1145/3534617`

[127] JASP Team. 2022. JASP [Computer software]. `https://jasp-stats.org/`

[128] Peraphan Jittrapirom, Valeria Caiati, Anna-Maria Feneri, Shima Ebrahimigharehbaghi, María J. Alonso-González, and Jishnu Narayan. 2017. Mobility as a Service: A Critical Review of Definitions, Assessments of Schemes, and Key Challenges. *Urban Planning* 2, 2 (2017), 13 – 25. `https://doi.org/10.17645/up.v2i2.931`

[129] Jeff Johnson. 2010. *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Rules*. Morgan Kaufmann Publishers, Burlington, MA, USA. `https://doi.org/10.1016/C2012-0-07128-1`

[130] Alita Joyce. 2019. Formative vs. Summative Evaluations. Retrieved 2022-10-22 from `https://www.nngroup.com/articles/formative-vs-summative-evaluations/`

[131] Juffrizal Karjanto, Nidzamuddin Md. Yusof, Jacques Terken, Frank Delbressine, Matthias Rauterberg, and Muhammad Zahir Hassan. 2018. Development of On-Road Automated Vehicle Simulator for Motion Sickness Studies. *International Journal of Driving Science* 1, 1 (2018), 1–12. `https://doi.org/10.5334/ijds.8`

[132] Kanwaldeep Kaur and Giselle Rampersad. 2018. Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars. *Journal of Engineering and Technology Management* 48 (2018), 87–96. `https://doi.org/10.1016/j.jengtecman.2018.04.006`

[133] Dagmar Kern and Albrecht Schmidt. 2009. Design Space for Driver-Based Automotive User Interfaces. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Essen, Germany) *(AutomotiveUI '09)*. Association for Computing Machinery, New York, NY, USA, 3–10. `https://doi.org/10.1145/1620509.1620511`

[134] Hyungil Kim and Joseph L. Gabbard. 2022. Assessing Distraction Potential of Augmented Reality Head-Up Displays for Vehicle Drivers. *Human Factors* 64, 5 (2022), 852–865. `https://doi.org/10.1177/0018720819844845`

[135] Sangwon Kim, Jennifer Jah Eun Chang, Hyun Ho Park, Seon Uk Song, Chang Bae Cha, Ji Won Kim, and Namwoo Kang. 2020. Autonomous Taxi Service Design and User Experience. *International Journal of Human-Computer Interaction* 36, 5 (2020), 429–448. `https://doi.org/10.1080/10447318.2019.1653556`

[136] René F. Kizilcec. 2016. How Much Information? Effects of Transparency on Trust in an Algorithmic Interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2390–2395. `https://doi.org/10.1145/2858036.2858402`

[137] Zoe Kleinman. 2022. Baidu unveils new self-driving taxi in China. Retrieved 2022-10-17 from `https://www.bbc.com/news/technology-62237612`

[138] Michael Kondzior. 2018. Akzeptanzskala – Methode zur Erfassung der Akzeptanz eines Systems. Deutsche Übersetzung. Retrieved 2018-05-25 from `https://www.hfes-europe.org/accept/accept_de.htm`

[139] Alexandra König, Christina Wirth, and Jan Grippenkoven. 2021. Generation Y's Information Needs Concerning Sharing Rides in Autonomous Mobility on Demand Systems. *Sustainability* 13, 14 (2021), 13 pages. `https://doi.org/10.3390/su13148095`

[140] Julian F.P. Kooij, Fabian Flohr, Ewoud A.I. Pool, and Dariu M. Gavrila. 2019. Context-Based Path Prediction for Targets with Switching Dynamics. *International Journal of Computer Vision* 127, 3 (2019), 239–262. `https://doi.org/10.1007/s11263-018-1104-4`

[141] Moritz Körber. 2019. Theoretical Considerations and Development of a Questionnaire to Measure Trust in Automation. In *Proceedings of the 20th Congress of the International Ergonomics Association* (Florence, Italy) *(IEA '18)*. Springer International Publishing, Cham, Switzerland, 13–30. `https://doi.org/10.1007/978-3-319-96074-6_2`

[142] Andreas Korthauer, Clemens Guenther, Andreas Hinrichs, Wen Ren, and Yiwen Yang. 2021. Watch Your Vehicle Driving at the City: Interior HMI with Augmented Reality for Automated Driving. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (Oldenburg, Germany) *(MobileHCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 59, 5 pages. `https://doi.org/10.1145/3406324.3425895`

[143] Christian Kray, Patrick Olivier, Amy Weihong Guo, Pushpendra Singh, Hai Nam Ha, and Phil Blythe. 2007. Taming Context: A Key Challenge in Evaluating the Usability of Ubiquitous Systems. Retrieved 2023-05-29 from `https://christian-kray.de/pub/KrayEtAl07_TamingContext.pdf`

[144] Fabian Kröger. 2016. Automated Driving in Its Social, Historical and Cultural Contexts. In *Autonomous Driving*. Springer, Berlin and Heidelberg, Germany, 41–68. `https://doi.org/10.1007/978-3-662-48847-8_3`

[145] Sven Krome, William Goddard, Stefan Greuter, Steffen P. Walz, and Ansgar Gerlicher. 2015. A Context-Based Design Process for Future Use Cases of Autonomous Driving: Prototyping AutoGym. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Nottingham, United Kingdom) *(AutomotiveUI '15)*. Association for Computing Machinery, New York, NY, USA, 265–272. `https://doi.org/10.1145/2799250.2799257`

[146] Rico Krueger, Taha H. Rashidi, and John M. Rose. 2016. Preferences for Shared Autonomous Vehicles. *Transportation Research Part C: Emerging Technologies* 69 (2016), 343–355. `https://doi.org/10.1016/j.trc.2016.06.015`

[147] Kenneth S. Kurani. 2019. *User Perceptions of Safety and Security: A Framework for a Transition to Electric-Shared-Automated Vehicles*. Technical Report. National Center for Sustainable Transportation, Davis, CA, USA. `https://doi.org/10.7922/G2891438`

[148] M. Kyriakidis, R. Happee, and J. C.F. de Winter. 2015. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour* 32 (2015), 127–140. `https://doi.org/10.1016/j.trf.2015.04.014`

[149] Andrew R. Lacher, Robert Grabowski, and Stephen Cook. 2014. Autonomy, Trust, and Transportation. In *Proceedings of the 2014 AAAI Spring Symposium (AAAI '14)* (Palo Alto, CA, USA). Association for the Advancement of Artificial Intelligence, Palo Alto, CA, USA, 42–49. `https://aaai.org/papers/07701-autonomy-trust-and-transportation/`

[150] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society* (Graz, Austria) *(USAB '08)*. Springer, Berlin and Heidelberg, Germany, 63–76. `https://doi.org/10.1007/978-3-540-89350-9_6`

[151] Patrícia S. Lavieri and Chandra R. Bhat. 2019. Modeling Individuals' Willingness to Share Trips with Strangers in an Autonomous Vehicle Future. *Transportation Research Part A: Policy and Practice* 124 (2019), 242–261. `https://doi.org/10.1016/j.tra.2019.03.009`

[152] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 1 (2004), 50–80. `https://doi.org/10.1518/hfes.46.1.50_30392`

[153] Michael W. Levin, Kara M. Kockelman, Stephen D. Boyles, and Tianxin Li. 2017. A general framework for modeling shared autonomous vehicles with dynamic network-loading and dynamic ride-sharing application. *Computers, Environment and Urban Systems* 64 (2017), 373–383. `https://doi.org/10.1016/j.compenvurbsys.2017.04.006`

[154] Youn-Kyung Lim, Erik Stolterman, and Josh Tenenberg. 2008. The Anatomy of Prototypes: Prototypes as Filters, Prototypes as Manifestations of Design Ideas. *ACM Trans. Comput.-Hum. Interact.* 15, 2, Article 7 (jul 2008), 27 pages. `https://doi.org/10.1145/1375761.1375762`

[155] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *Proceedings of the 2014 European Conference on Computer Vision* (Zurich, Switzerland) *(ECCV '14)*, Fleet David, Tomas Pajdla, Schiele Bernt, and Tuytelaars Tinne (Eds.). Springer International Publishing, Cham, Switzerland, 740–755. `https://doi.org/10.1007/978-3-319-10602-1_48`

[156] Patrick Lindemann, Tae-Young Lee, and Gerhard Rigoll. 2018. Catch my drift: Elevating situation awareness for highly automated driving with an explanatory windshield display user interface. *Multimodal Technologies and Interaction* 2, 4 (2018), 27 pages. `https://doi.org/10.3390/mti2040071`

[157] Todd Litman. 2023. Autonomous Vehicle Implementation Predictions. Retrieved 2023-04-15 from `https://www.vtpi.org/avip.pdf`

[158] Martin Ludvigsen. 2006. *Designing for Social Interaction Physical, Co-located Social Computing*. Dissertation. Aarhus School of Architecture Denmark. `http://www.mvl.dk/dissertation/DesigningforSocialInteraction_dissertationMVL_web.pdf`

[159] Martin Lundquist. 2018. *Autonomous Bus Passenger Experience*. Master's thesis. Umeå University, Faculty of Science and Technology, Department of Applied Physics and Electronics, Umeå, Sweden. `https://umu.diva-portal.org/smash/record.jsf?pid=diva2%3A1218637&dswid=-6757`

[160] Ana Mackay, Inês Fortes, Catarina Santos, Dário Machado, Patrícia Barbosa, Vera Vilas Boas, João Pedro Ferreira, Nélson Costa, Carlos Silva, and Emanuel Sousa. 2019. The Impact of Autonomous Vehicles' Active Feedback on Trust. In *Proceedings of the International Conference on Applied Human Factors and Ergonomics: Advances in Safety Management and Human Factors* (Washington, DC, USA) *(AHFE '19)*. Springer International Publishing, Cham, Switzerland, 342–352. `https://doi.org/10.1007/978-3-030-20497-6_32`

[161] Margarita Martínez-Díaz and Francesc Soriguera. 2018. Autonomous vehicles: theoretical and practical challenges. *Transportation Research Procedia* 33 (2018), 275–282. `https://doi.org/10.1016/j.trpro.2018.10.103`

[162] Kohei Matsumura and David S. Kirk. 2018. On Active Passengering: Supporting In-Car Experiences. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 154 (jan 2018), 23 pages. `https://doi.org/10.1145/3161176`

[163] Philipp Mayring. 2010. Handbuch Qualitative Forschung in der Psychologie. In *Handbuch Qualitative Forschung in der Psychologie*, Günter Mey and Katja Mruck (Eds.). VS Verlag für Sozialwissenschaften & Springer Fachmedien, Wiesbaden, Germany, 601–613. `https://doi.org/10.1007/978-3-531-92052-8_42`

[164] Michael McTear, Zoraida Callejas, and David Griol. 2016. *The Conversational Interface: Talking to Smart Devices*. Springer, Cham, Switzerland. 422 pages. `https://doi.org/10.1007/978-3-319-32967-3`

[165] Michael F. McTear. 2017. The Rise of the Conversational Interface: A New Kid on the Block? In *Future and Emerging Trends in Language Technology – Machine Learning and Big Data*, José F. Quesada, Francisco J. Martín Mateos, and Teresa López-Soto (Eds.). Springer, Cham, Switzerland, 38–49. `https://doi.org/10.1007/978-3-319-69365-1`

[166] Mercedes-Benz AG. 2023. Mercedes-Benz Drive Pilot. Retrieved 2023-02-19 from `https://www.mercedes-benz.de/passengercars/technology-innovation/mercedes-benz-drive-pilot/stage.module.html`

[167] Johanna Meurer, Christina Pakusch, Gunnar Stevens, Dave Randall, and Volker Wulf. 2020. A Wizard of Oz Study on Passengers' Experiences of a Robo-Taxi Service in Real-Life Settings. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1365–1377. `https://doi.org/10.1145/3357236.3395465`

[168] Paul Milgram and Fumio Kishino. 1994. A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information Systems* E77-D, 12 (1994), 1321–1329. `https://search.ieice.org/bin/summary.php?id=e77-d_12_1321`

[169] Alexander G. Mirnig, Vivien Wallner, Magdalena Gärtner, Alexander Meschtscherjakov, and Manfred Tscheligi. 2020. Capacity Management in an Automated Shuttle Bus: Findings from a Lab Study. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Virtual Event, DC, USA) *(AutomotiveUI '20)*. Association for Computing Machinery, New York, NY, USA, 270–279. `https://doi.org/10.1145/3409120.3410665`

[170] Alex Mohebbi. 2018. Why Conversational Interfaces are taking us back to the Dark Ages of Usability. Retrieved 2020-09-20 from `https://tinyurl.com/4zh6fw73`

[171] Mareike Möhlmann. 2015. Collaborative consumption: determinants of satisfaction and the likelihood of using a sharing economy option again. *Journal of Consumer Behaviour* 14, 3 (2015), 193–207. `https://doi.org/10.1002/cb.1512`

[172] Kate Moran. 2016. The Four Dimensions of Tone of Voice. Retrieved 2023-03-19 from `https://www.nngroup.com/articles/tone-of-voice-dimensions/`

[173] Alyona Morozova, Verena Rheinstädter, and Dieter Wallach. 2019. MixedUX: A Mixed Prototyping Framework for Usability Testing in Augmented Reality. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion* (San Diego, CA, USA) *(DIS '19 Companion)*. Association for Computing Machinery, New York, NY, USA, 41–44. `https://doi.org/10.1145/3301019.3325146`

[174] Bonnie M. Muir. 1994. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics* 37, 11 (1994), 1905–1922. `https://doi.org/10.1080/00140139408964957`

[175] Andrea Isabell Müller, Veronika Weinbeer, and Klaus Bengler. 2019. Using the Wizard of Oz Paradigm to Prototype Automated Vehicles: Methodological Challenges. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 181–186. `https://doi.org/10.1145/3349263.3351526`

[176] Jakob Nielsen. 1994. 10 Usability Heuristics for User Interface Design. Retrieved 2017-03-10 from `https://www.nngroup.com/articles/ten-usability-heuristics/`

[177] Jakob Nielsen. 1994. Heuristic Evaluation. In *Usability Inspection Methods* (1 ed.), Jakob Nielsen and Robert L. Mack (Eds.). John Wiley & Sons, Hoboken, NJ, USA, Chapter 2, 25–62. `https://www.nngroup.com/books/usability-inspection-methods/`

[178] Jakob Nielsen. 1994. Usability Inspection Methods. In *Conference Companion on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '94)*. Association for Computing Machinery, New York, NY, USA, 413–414. `https://doi.org/10.1145/259963.260531`

[179] Nils J. Nilsson. 1998. *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann Publishers, San Francisco, CA. `https://doi.org/10.1016/b978-0-08-049945-1.50001-0`

[180] Amelie Nolte, Karolin Lueneburg, Dieter P. Wallach, and Nicole Jochems. 2022. Creating Personas for Signing User Populations: An Ability-Based Approach to User Modelling in HCI. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility* (Athens, Greece) *(ASSETS '22)*. Association for Computing Machinery, New York, NY, USA, Article 50, 6 pages. `https://doi.org/10.1145/3517428.3550364`

[181] Sina Nordhoff, Joost de Winter, Ruth Madigan, Natasha Merat, Bart van Arem, and Riender Happee. 2018. User acceptance of automated shuttles in Berlin-Schöneberg: A questionnaire study. *Transportation Research Part F: Traffic Psychology and Behaviour* 58 (2018), 843–854. `https://doi.org/10.1016/j.trf.2018.06.024`

[182] Sina Nordhoff, Joost de Winter, William Payre, Bart Van Arem, and Riender Happee. 2019. What Impressions Do Users Have After a Ride in an Automated Shuttle? An Interview Study. *Transportation Research Part F: Traffic Psychology and Behaviour* 63 (2019), 252–269. `https://doi.org/10.1016/j.trf.2019.04.009`

[183] Donald A. Norman. 1990. The "Problem" with Automation: Inappropriate Feedback and Interaction, Not "Overautomation". *Philosophical Transactions of The Royal Society B, Biological Sciences* 327, 1241 (1990), 585–593. `https://doi.org/10.1098/rstb.1990.0101`

[184] William Odom, John Zimmerman, Scott Davidoff, Jodi Forlizzi, Anind K. Dey, and Min Kyung Lee. 2012. A Fieldwork of the Future with User Enactments. In *Proceedings of the Designing Interactive Systems Conference* (Newcastle Upon Tyne, United Kingdom) *(DIS '12)*. Association for Computing Machinery, New York, NY, USA, 338–347. `https://doi.org/10.1145/2317956.2318008`

[185] Luis Oliveira, Christopher Burns, Jacob Luton, Sumeet Iyer, and Stewart Birrell. 2020. The influence of system transparency on trust: Evaluating interfaces in a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour* 72 (2020), 280–296. `https://doi.org/10.1016/j.trf.2020.06.001`

[186] Morin Ostkamp and Christian Kray. 2014. Supporting Design, Prototyping, and Evaluation of Public Display Systems. In *Proceedings of the 2014 ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (Rome, Italy) *(EICS '14)*. Association for Computing Machinery, New York, NY, USA, 263–272. `https://doi.org/10.1145/2607023.2607035`

[187] Ganesh Pai, Sarah Widrow, Jaydeep Radadiya, Cole D. Fitzpatrick, Michael Knodler, and Anuj K. Pradhan. 2020. A Wizard-of-Oz experimental approach to study the human factors of automated vehicles: Platform and methods evaluation. *Traffic Injury Prevention* 21, S1 (2020), S140–S144. `https://doi.org/10.1080/15389588.2020.1810243`

[188] Christina Pakusch, Gunnar Stevens, Alexander Boden, and Paul Bossauer. 2018. Unintended effects of autonomous driving: A study on mobility preferences in the future. *Sustainability* 10, 7 (2018), 22 pages. `https://doi.org/10.3390/su10072404`

[189] Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors* 39, 2 (1997), 230–253. `https://doi.org/10.1518/001872097778543886`

[190] Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens. 2000. A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans.* 30, 3 (2000), 286–297. `https://doi.org/10.1109/3468.844354`

[191] Marco Pavone. 2016. Autonomous Mobility-on-Demand Systems for Future Urban Mobility. In *Autonomous Driving – Technical, Legal and Social Aspects*, Markus Mauerer, J. Christian Gerdes, Barbara Lenz, and Hermann Winner (Eds.). Springer, Berlin and Heidelberg, Germany, Chapter 19, 387–404. `https://doi.org/10.1007/978-3-662-48847-8_19`

[192] Ralf Philipsen, Teresa Brell, Hannah Biermann, and Martina Ziefle. 2020. On the Road Again - Explanatory Factors for the Users' Willingness to Replace Private Cars by Autonomous on-Demand Shuttle Services. In *Proceedings of the 2020 International Conference on Applied Human Factors and Ergonomics: Advances in Human Aspects of Transportation* (Virtual, USA) *(AHFE '20)*. Springer International Publishing, Cham, Switzerland, 173–185. `https://doi.org/10.1007/978-3-030-50943-9_23`

[193] Ralf Philipsen, Teresa Brell, and Martina Ziefle. 2018. Carriage without a Driver – User Requirements for Intelligent Autonomous Mobility Services. In *Proceedings of the International Conference on Applied Human Factors and Ergonomics: Advances in Human Aspects of Transportation* (Orlando, FL, USA) *(AHFE '18)*. Springer International Publishing, Cham, Switzerland, 339–350. `https://doi.org/10.1007/978-3-319-93885-1_31`

[194] Jinan Piao, Mike McDonald, Nick Hounsell, Matthieu Graindorge, Tatiana Graindorge, and Nicolas Malhene. 2016. Public Views towards Implementation of Automated Vehicles in Urban Areas. *Transportation Research Procedia* 14 (2016), 2168–2177. `https://doi.org/10.1016/j.trpro.2016.05.232`

[195] Caroline Pigeon, Aline Alauzet, and Laurence Paire-Ficout. 2021. Factors of acceptability, acceptance and usage for non-rail autonomous public transport vehicles: a systematic literature review. *Transportation Research Part F: Traffic Psychology and Behaviour* 81 (aug 2021), 251–270. `https://doi.org/10.1016/j.trf.2021.06.008`

[196] Amalia Polydoropoulou, Ioannis Tsouros, Nikolas Thomopoulos, Cristina Pronello, Arnór Elvarsson, Haraldur Sigþórsson, Nima Dadashzadeh, Kristina Stojmenova, Jaka Sodnik, Stelios Neophytou, Domokos Esztergár-Kiss, Jamil Hamadneh, Graham Parkhurst, Shelly Etzioni, Yoram Shiftan, and Floridea Di Ciommo. 2021. Who Is Willing to Share Their AV? Insights about Gender Differences among Seven Countries. *Sustainability* 13, 9 (2021), 19 pages. `https://doi.org/10.3390/su13094769`

[197] Ryan Powell. 2019. Trusting Driverless Cars – How Waymo designed an experience that reassures riders every step of the way. Retrieved 2021-09-02 from `https://design.google/library/trusting-driverless-cars/`

[198] Beatrice Rammstedt. 2010. Reliabilität, Validität, Objektivität. In *Handbuch der sozialwissenschaftlichen Datenanalyse* (1 ed.), Christof Wolf and Henning Best (Eds.). VS Verlag für Sozialwissenschaften & Springer Fachmedien, Wiesbaden, Germany, Chapter 11, 239 – 258. `https://doi.org/10.1007/978-3-531-92038-2_11`

[199] Beatrice Rammstedt and Oliver P. John. 2007. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality* 41, 1 (2007), 203–212. `https://doi.org/10.1016/j.jrp.2006.02.001`

[200] Beatrice Rammstedt, Christoph J. Kemper, Mira Céline Klein, Constanze Beierlein, and Anastassiya Kovaleva. 2013. A Short Scale for Assessing the Big Five Dimensions of Personality: 10 Item Big Five Inventory (BFI-10). *methods, data, analyses* 7, 2 (2013), 5–32. `https://doi.org/10.12758/mda.2013.013`

[201] Manuel Alector Ribeiro, Dogan Gursoy, and Oscar Hengxuan Chi. 2022. Customer Acceptance of Autonomous Vehicles in Travel and Tourism. *Journal of Travel Research* 61, 3 (2022), 620–636. `https://doi.org/10.1177/0047287521993578`

[202] Natalie Tara Richardson. 2020. *Konzeption und Langzeittest der Mensch-Maschine-Schnittstelle für das hochautomatisierte Fahren im Lkw*. Dissertation. Technical University of Munich, Munich, Germany. `https://mediatum.ub.tum.de/1520460`

[203] Natalie T. Richardson, Lukas Flohr, and Britta Michel. 2018. Takeover Requests in Highly Automated Truck Driving: How Do the Amount and Type of Additional Information Influence the Driver–Automation Interaction? *Multimodal Technologies and Interaction* 2, 4 (2018), 14 pages. `https://doi.org/10.3390/mti2040068`

[204] Mark O. Riedl. 2019. Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies* 1, 1 (feb 2019), 33–36. `https://doi.org/10.1002/hbe2.117`

[205] Joanna Roberts. 2015. Driverless taxis to become a major form of transport 'in 10 years'. *Horizon – The EU Research and Innovation Magazine* (2015). Retrieved 2023-04-16 from `https://ec.europa.eu/research-and-innovation/en/horizon-magazine/driverless-taxis-become-major-form-transport-10-years`

[206] José Rodríguez. 2022. Cruise Self-Driving Taxis Start Daytime Rides in San Francisco. Retrieved 2023-01-02 from `https://tinyurl.com/2tfhehxk`

[207] RStudio Team. 2021. Integrated Development for R [Computer Software]. `http://www.rstudio.com/`

[208] Jim Rudd, Ken Stern, and Scott Isensee. 1996. Low vs. High-Fidelity Prototyping Debate. *Interactions* 3, 1 (1996), 76–85. `https://doi.org/10.1145/223500.223514`

[209] James A. Russell, Anna Weiss, and Gerald A. Mendelsohn. 1989. Affect Grid: A Single-Item Scale of Pleasure and Arousal. *Journal of Personality and Social Psychology* 57, 3 (1989), 493–502. `https://doi.org/10.1037/0022-3514.57.3.493`

[210] SAE International and ISO. 2021. J3016: Surface Vehicle Recommended Practice: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. `https://doi.org/10.4271/J3016_202104`

[211] Arto O. Salonen. 2018. Passenger's subjective traffic safety, in-vehicle security and emergency management in the driverless shuttle bus in Finland. *Transport Policy* 61 (2018), 106–110. `https://doi.org/10.1016/j.tranpol.2017.10.011`

[212] Arto O. Salonen and Noora Haavisto. 2019. Towards Autonomous Transportation. Passengers' Experiences, Perceptions and Feelings in a Driverless Shuttle Bus in Finland. *Sustainability* 11, 3 (2019), 19 pages. `https://doi.org/10.3390/su11030588`

[213] Angela Sanguinetti, Ken Kurani, and Beth Ferguson. 2019. *Is it OK to Get in a Car with a Stranger? Risks and Benefits of Ride-pooling in Shared Automated Vehicles*. Technical Report. Institute of Transportation Studies, University of California, Davis, CA, USA. `https://doi.org/10.7922/G27M064Z`

[214] Javier Morales Sarriera, Germán Escovar Álvarez, Kelly Blynn, Andrew Alesbury, Timothy Scully, and Jinhua Zhao. 2017. To Share or Not to Share: Investigating the Social Aspects of Dynamic Ridesharing. *Transportation Research Record* 2605, 1 (2017), 109–117. `https://doi.org/10.3141/2605-11`

[215] Nadav Savio and Jared Braiterman. 2007. Design Sketch: The Context of Mobile Interaction. Retrieved 2023-03-25 from `https://www.giantant.com/publications/mobile_context_model.pdf`

[216] Anna Schieben, Marc Wilbrink, Carmen Kettwich, Ruth Madigan, Tyron Louw, and Natasha Merat. 2019. Designing the interaction of automated vehicles with other traffic participants: design considerations based on human needs and expectations. *Cognition, Technology, and Work* 21 (2019), 69–85. `https://doi.org/10.1007/s10111-018-0521-z`

[217] Peter Schlanstein. 2010. Verbesserung der Grundlagen und Methodik in der Unfallanalyse. In *Deutscher Verkehrsexpertentag 2010* (Münster, Germany). Hochschule für Polizei und öffentliche Verwaltung Nordrhein-Westfalen, Gelsenkirchen, Germany, 5 pages. Retrieved 2023-04-16 from `https://www.hspv.nrw.de/fileadmin/Newsletter/Artikel_Schlanstein.pdf`

[218] Albrecht Schmidt. 2000. Implicit Human Computer Interaction Through Context. *Personal Technologies* 4 (2000), 191–199. `https://doi.org/10.1007/BF01324126`

[219] Hans-Peter Schöner and Bernhard Morys. 2015. Dynamische Fahrsimulatoren. In *Handbuch Fahrerassistenzsysteme* (3 ed.), H. Winner, S. Hakuli, F. Lotz, and C. Singer (Eds.). Springer Vieweg, Wiesbaden, Germany, Chapter 9, 139–155. `https://doi.org/10.1007/978-3-658-05734-3_9`

[220] Martin Schrepp. 2018. User Experience Questionnaire Handbook. `https://doi.org/10.13140/rg.2.1.2815.0245`

[221] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 4 (2017), 40–44. `https://doi.org/10.9781/ijimai.2017.445`

[222] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 6 (2017), 103. `https://doi.org/10.9781/ijimai.2017.09.001`

[223] Martin Schrepp and Jörg Thomaschewski. 2019. Design and Validation of a Framework for the Creation of User Experience Questionnaires. *International Journal of Interactive Multimedia and Artificial Intelligence* 5, 7 (2019), 88–95. `https://doi.org/10.9781/ijimai.2019.06.006`

[224] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments* 10, 3 (2001), 266–281. `https://doi.org/10.1162/105474601300343603`

[225] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2016. Igroup Presence Questionnaire (IPQ). Retrieved 2019-06-23 from `http://www.igroup.org/pq/ipq/index.php`

[226] Martina Schuß, Carina Manger, Andreas Löcken, and Andreas Riener. 2022. You'Ll Never Ride Alone: Insights into Women's Security Needs in Shared Automated Vehicles. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Seoul, Republic of Korea) *(AutomotiveUI '22)*. Association for Computing Machinery, New York, NY, USA, 13–23. `https://doi.org/10.1145/3543174.3546848`

[227] Martina Schuß, Philipp Wintersberger, and Andreas Riener. 2021. Let's Share a Ride into the Future: A Qualitative Study Comparing Hypothetical Implementation Scenarios of Automated Vehicles. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 156, 11 pages. `https://doi.org/10.1145/3411764.3445609`

[228] Martina Schuß, Philipp Wintersberger, and Andreas Riener. 2021. Security Issues in Shared Automated Mobility Systems: A Feminist HCI Perspective. *Multimodal Technologies and Interaction* 5, 8 (2021), 16 pages. `https://doi.org/10.3390/mti5080043`

[229] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using Presence Questionnaires in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. `https://doi.org/10.1145/3290605.3300590`

[230] Susan Shaheen. 2018. Shared Mobility: The Potential of Ridehailing and Pooling. In *Three Revolutions : Steering Automated, Shared, and Electric Vehicles to a Better Future*. Island Press, Washington, DC, USA, Chapter 3, 55–75. `https://doi.org/10.5822/978-1-61091-906-7_3`

[231] William R. Sherman and Alan B. Craig. 2003. *Understanding Virtual Reality: Interface, Application, and Design*. Morgan Kaufmann, Burlington, MA, USA. `https://www.sciencedirect.com/book/9781558603530/understanding-virtual-reality`

[232] Amir Shevat. 2017. *Designing Bots: Creating Conversational Experiences*. O'Reilley Media, Sebastopol, CA, USA. `https://www.oreilly.com/library/view/designing-bots/9781491974810/`

[233] Ben Shneiderman. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human-Computer Interaction* 36, 6 (mar 2020), 495–504. `https://doi.org/10.1080/10447318.2020.1741118`

[234] Ben Shneiderman. 2020. Human-Centered Artificial Intelligence: Three Fresh Ideas. *AIS Transactions on Human-Computer Interaction* 12, 3 (2020), 109–124. `https://doi.org/10.17705/1thci.00131`

[235] Marion K. Slack and Jolaine R. Draugalis. 2001. Establishing the internal and external validity of experimental studies. *American Journal of Health-System Pharmacy* 58, 22 (2001), 2173–2181. `https://doi.org/10.1093/ajhp/58.22.2173`

[236] Hannah Snyder. 2019. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research* 104 (2019), 333–339. `https://doi.org/10.1016/j.jbusres.2019.07.039`

[237] Charles D. Spielberger. 1983. State-Trait Anxiety Inventory for Adults (STAI-AD). *APA PsycTests* (1983). `https://doi.org/10.1037/t06496-000`

[238] Kevin Spieser, Kyle Treleaven, Rick Zhang, Emilio Frazzoli, Daniel Morton, and Marco Pavone. 2014. Toward a Systematic Approach to the Design and Evaluation of Automated Mobility-on-Demand Systems: A Case Study in Singapore. In *Road Vehicle Automation*, Gereon Meyer and Sven Beiker (Eds.). Springer International Publishing, Cham, Switzerland, 229–245. `https://doi.org/10.1007/978-3-319-05990-7_20`

[239] Annika Stampf, Mark Colley, and Enrico Rukzio. 2022. Towards Implicit Interaction in Highly Automated Vehicles - A Systematic Literature Review. *Proc. ACM Hum.-Comput. Interact.* 6, MHCI, Article 191 (sep 2022), 21 pages. `https://doi.org/10.1145/3546726`

[240] Kerstin Stark, Kay Gade, and Dirk Heinrichs. 2019. What Does the Future of Automated Driving Mean for Public Transportation? *Transportation Research Record* 2673, 2 (2019), 85–93. `https://doi.org/10.1177/0361198119827578`

[241] Statistisches Bundesamt (Destatis). 2022. Verkehrsunfälle - Zeitreihen 2021. Retrieved 2023-04-16 from `https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/_inhalt.html`

[242] Toni Steimle and Dieter Wallach. 2018. *Collaborative UX Design*. dpunkt.verlag, Heidelberg, Germany. 240 pages. `http://www.collaborative-uxdesign.com/`

[243] Marc Stickdorn and Jakob Schneider. 2017. *This is Service Design Thinking* (7 ed.). BIS Publishers, Amsterdam, Netherlands. `http://thisisservicedesignthinking.com/`

[244] Kyoko Sugisaki and Andreas Bleiker. 2020. GermanUPA Webinar: Natural Language Conversation als UI - Traum oder Albtraum? Retrieved 2020-03-30 from `https://germanupa.de/berufsverband-german-upa/aktuelles/webinar-natural-language-conversation-als-ui-traum-albtraum`

[245] Kyoko Sugisaki and Andreas Bleiker. 2020. Usability Guidelines and Evaluation Criteria for Conversational User Interfaces: A Heuristic and Linguistic Approach. In *Proceedings of Mensch Und Computer 2020* (Magdeburg, Germany) *(MuC '20)*. Association for Computing Machinery, New York, NY, USA, 309–319. `https://doi.org/10.1145/3404983.3405505`

[246] Tesla. 2022. Autopilot | Model 3 and Model Y. Retrieved 2022-11-11 from `https://www.youtube.com/watch?v=-FeMwPUAOLM`

[247] Martin Thaler. 2018. Our Faculty | Martin Thaler on Prototyping. Retrieved 2022-10-15 from `https://www.youtube.com/watch?v=O1MQd1-xTDg`

[248] The Guardian. 2022. California allows driverless taxi service to operate in San Francisco. Retrieved 2022-10-17 from `https://www.theguardian.com/technology/2022/jun/03/california-driverless-taxi-cars-san-francisco`

[249] The jamovi project. 2022. jamovi [Computer Software]. `https://www.jamovi.org`

[250] The Waymo Team. 2023. Keeping riders and other road users safe with Safe Exit features. Retrieved 2023-08-02 from `https://waymo.com/blog/2023/05/keeping-riders-and-other-road-users.html`

[251] Nikolas Thomopoulos and Moshe Givoni. 2015. The autonomous car—a blessing or a curse for the future of low carbon mobility? An exploration of likely vs. desirable outcomes. *European Journal of Futures Research* 3, 14 (2015), 14 pages. `https://doi.org/10.1007/s40309-015-0071-z`

[252] Alexander Toet, Daisuke Kaneko, Shota Ushiama, Sofie Hoving, Inge de Kruijf, Anne-Marie Brouwer, Victor Kallen, and Jan B. F. van Erp. 2018. EmojiGrid: A 2D Pictorial Scale forthe Assessment of Food Elicited Emotions. *Frontiers in Psychology* 9, 2396 (2018), 21 pages. `https://doi.org/10.3389/fpsyg.2018.02396`

[253] Munesh Chandra Trivedi and Mohammadi Akheela Khanum. 2012. Role Of Context In Usability Evaluations: A Review. *Advanced Computing: An International Journal* 3, 2 (2012), 69–78. `https://doi.org/10.5121/acij.2012.3208`

[254] United Nations. 2022. The Sustainable Development Goals Report 2022. `https://unstats.un.org/sdgs/report/2022/The-Sustainable-Development-Goals-Report-2022.pdf`

[255] Stefano Valtolina, Barbara Rita Barricelli, and Serena Di Gaetano. 2020. Communicability of traditional interfaces VS chatbots in healthcare and smart home domains. *Behaviour and Information Technology* 39, 1 (2020), 108–132. `https://doi.org/10.1080/0144929X.2019.1637025`

[256] Jinke D. Van der Laan, Adriaan Heino, and Dick De Waard. 1997. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies* 5, 1 (feb 1997), 1–10. `https://doi.org/10.1016/S0968-090X(96)00025-3`

[257] Gustavo Velasco-Hernandez, De Jong Yeong, John Barry, and Joseph Walsh. 2020. Autonomous Driving Architectures, Perception and Data Fusion: A Review. In *Proceedings of the 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP '20)* (Cluj-Napoca, Romania). Institute of Electrical and Electronics Engineers, New York, NY, USA, 315–321. `https://doi.org/10.1109/ICCP51029.2020.9266268`

[258] Viswanath Venkatesh and Fred D. Davis. 2000. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science* 46, 2 (2000), 186–204. `https://doi.org/10.1287/mnsc.46.2.186.11926`

[259] Robert A. Virzi, Jeffrey L. Sokolov, and Demetrios Karis. 1996. Usability Problem Identification Using Both Low- and High-Fidelity Prototypes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, British Columbia, Canada) *(CHI '96)*. Association for Computing Machinery, New York, NY, USA, 236–243. `https://doi.org/10.1145/238386.238516`

[260] Dieter P. Wallach, Lukas A. Flohr, and Annika Kaltenhauser. 2020. Beyond the Buzzwords: On the Perspective of AI in UX and Vice Versa. In *Proceedings of the 1st International Conference on Artificial Intelligence in HCI, Held as Part of the 22nd HCI International Conference on Human-Computer Interaction* (Copenhagen, Denmark) *(HCII '20)*. Springer International Publishing, Cham, Switzerland, 146–166. `https://doi.org/10.1007/978-3-030-50334-5_10`

[261] Peter Wang, Srinath Sibi, Brian Mok, and Wendy Ju. 2017. Marionette: Enabling On-Road Wizard-of-Oz Autonomous Driving Studies. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna, Austria) *(HRI '17)*. Association for Computing Machinery, New York, NY, USA, 234–243. `https://doi.org/10.1145/2909824.3020256`

[262] Todd Zaki Warfel. 2009. *Prototyping – A Practitioner's Guide*. Rosenfeld Media, Brooklyn, NY, USA. `https://rosenfeldmedia.com/books/prototyping/`

[263] Paul Watzlawick, Janet Beavin Bavelas, and Don D. Jackson. 2011. *Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies, and Paradoxes*. Norton Professional Books, New York, NY, USA. `https://wwnorton.com/books/9780393710595`

[264] Waymo. 2022. Taking a rider-only trip with the Waymo Driver in San Francisco. Retrieved 2022-11-11 from `https://www.youtube.com/watch?v=ospoTAyEdDQ`

[265] Waymo. 2022. Waymo One. Retrieved 2022-10-17 from `https://waymo.com/waymo-one/`

[266] Waypoint - The official Waymo blog. 2022. Wheels up for Waymo as we expand our 24/7 rider-only territories. Retrieved 2023-01-02 from `https://blog.waymo.com/2022/12/wheels-up-for-waymo-as-we-expand.html`

[267] Arno Weber. 2023. Safety und Security – unterschiedlich und doch untrennbar / 1 Unterschiede und Gemeinsamkeiten von Safety und Security. Retrieved 2023-02-20 from `https://www.haufe.de/arbeitsschutz/arbeitsschutz-office-professional/safety-und-security-unterschiedlich-und-doch-untrennbar-1-unterschiede-und-gemeinsamkeiten-von-safety-und-security_idesk_PI13633_HI14780245.html`

[268] Daniel Wessel, Christiane Attig, and Thomas Franke. 2019. ATI-S - An Ultra-Short Scale for Assessing Affinity for Technology Interaction in User Studies. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC'19)*. Association for Computing Machinery, New York, NY, USA, 147–154. `https://doi.org/10.1145/3340764.3340766`

[269] Christopher D. Wickens, John D. Lee, Yili Liu, and Sallie E. Gordon-Becker. 2004. Automation. In *An Introduction to Human Factors Engineering* (2 ed.). Pearson Education, New York, NY, USA, 398–415. `https://www.pearson.com/store/p/introduction-to-human-factors-engineering-pearson-new-international-edition/P200000005734/9781292035512`

[270] Philipp Wintersberger, Anna-Katharina Frison, Andreas Riener, and Tamara von Sawitzky. 2018. Fostering User Acceptance and Trust in Fully Automated Vehicles: Evaluating the Potential of Augmented Reality. *Presence: Teleoperators and Virtual Environments* 27, 1 (2018), 46–62. `https://doi.org/10.1162/pres_a_00320`

[271] Philipp Wintersberger, Tamara von Sawitzky, Anna-Katharina Frison, and Andreas Riener. 2017. Traffic Augmentation as a Means to Increase Trust in Automated Driving Systems. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter* (Cagliari, Italy) *(CHItaly '17)*. Association for Computing Machinery, New York, NY, USA, Article 17, 7 pages. `https://doi.org/10.1145/3125571.3125600`

[272] Ingo Wolf. 2015. Wechselwirkung Mensch und autonomer Agent. In *Autonomes Fahren*, M. Maurer, J.C. Gerdes, B. Lenz, and H. Winner (Eds.). Springer Vieweg, Berlin and Heidelberg, Germany, Chapter 6, 103–125. `https://doi.org/10.1007/978-3-662-45854-9_6`

[273] World Health Organization. 2018. Global Status Report on Road Safety 2018. Retrieved 2023-03-03 from `https://www.who.int/publications/i/item/9789241565684`

[274] World Health Organization. 2022. Road traffic injuries. Retrieved 2022-10-17 from `https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries`

[275] World Medical Association. 2000. World Medical Association Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. *JAMA* 310, 20 (2000), 2191–2194. `https://doi.org/10.1001/jama.2013.281053`

[276] Menno D. Yap, Gonçalo Correia, and Bart van Arem. 2016. Preferences of travellers for using automated vehicles as last mile public transport of multimodal train trips. *Transportation Research Part A: Policy and Practice* 94 (2016), 1–16. `https://doi.org/10.1016/j.tra.2016.09.003`

[277] De Jong Yeong, Gustavo Velasco-Hernandez, John Barry, and Joseph Walsh. 2021. Sensor and Sensor Fusion Technology in Autonomous Vehicles: A Review. *Sensors* 21, 6 (2021), 1–37. `https://doi.org/10.3390/s21062140`

[278] Larissa Zacherl, Jonas Radlmayr, and Klaus Bengler. 2020. Constructing a Mental Model of Automation Levels in the Area of Vehicle Guidance. In *Proceedings of the 3rd International Conference on Intelligent Human Systems Integration* (Modena, Italy) *(IHSI '20)*. Springer International Publishing, Cham, Switzerland, 73–79. `https://doi.org/10.1007/978-3-030-39512-4_12`

[279] Michelle X. Zhou, Gloria Mark, Jingyi Li, and Huahai Yang. 2019. Trusting Virtual Agents: The Effect of Personality. *ACM Trans. Interact. Intell. Syst.* 9, 2–3, Article 10 (mar 2019), 36 pages. `https://doi.org/10.1145/3232077`

[280] David Zipper. 2022. Self-Driving Taxis Are Causing All Kinds of Trouble in San Francisco. Retrieved 2023-01-02 from `https://slate.com/technology/2022/12/san-francisco-waymo-cruise-self-driving-cars-robotaxis.html`

Autonomous vehicles (AVs; SAE levels 4 and 5) face substantial challenges regarding acceptance, human factors, and user experience. Human-machine interfaces (HMIs) offer the potential to account for those and facilitate AV adoption.

Since AVs' capabilities and availability are still limited, suitable prototyping methods are required to create, evaluate, and optimize novel HMI concepts from early development phases. In all human-centered design activities, physical and social contexts are vital. This thesis argues for applying context-based interface prototyping of human-AV interactions to account for their interrelation with contextual factors.

We adopt a 'research in and through design' approach and explore the two intertwined areas: design and prototyping. Regarding the latter, we concentrate on straightforward methods. We demonstrate an immersive video-based approach for lab simulation of AVs and a wizard-of-oz-based method for on-road AV simulation and prototyping of HMIs providing real-time information. We apply these methods in empirical studies to assess their suitability and explore HMI concepts created to counter the aforementioned challenges. Thereby, we investigate the potential of (AR-based) object detection visualization and concepts for mobile and in-vehicle interaction with (shared) AVs.

Based on the findings, we provide design and prototyping recommendations that will aid researchers and practitioners in creating suitable human-AV interactions.

SIC Saarland Informatics Campus

UMTL

ERGOSIGN