Towards user friendly autonomous on-demand mobility: insights from a Wizard of Oz pretest

Fabian Schlichtherle[1], Sven Hornburg[2], Mario Rieker[2], & Wolfram Remlinger[1]
[1]University of Stuttgart, [2]University of Applied Sciences Schwäbisch Gmünd
Germany

Abstract

Wizard of Oz (WoOz) experiments are a good and well established means of testing technologies whose practical applicability has not yet progressed far enough. This also applies to the field of highly automated vehicles where driver vehicle interaction in passenger cars has been especially the focus of extensive attention. One aspect of WoOz research in the automobile context that has received much less awareness to date is shared automated vehicles (SAVs) as part of ride-pooling services. Findings of previous research on human vehicle interaction in WoOz environments have provided valuable insights. However, fundamental differences in the interior layouts as well as use cases and user groups of both passenger cars and SAVs, make it necessary to extend the application of this research method. We report on a preliminary test in which we realized the entire usage process of a ride with an autonomous ride-pooling service including multi-passenger interaction as a WoOz experiment. The main emphasis was on questions regarding technical feasibility as well as interface design aspects. Results indicated that WoOz is a suitable method to investigate user needs in SAVs and that this technique is a useful addition to the user-centred design approach.

Introduction

Necessity and current status of SAVs

High volumes of traffic, especially from private vehicles, are increasingly pushing cities to the limits of their infrastructure. In this context, automated driving can function as a key technology to make transportation of people as well as goods more efficient. However, vehicle automation alone will not solve the capacity problems of urban transport. On the contrary, it could even lead to an increase of car usage in the modal split, hence emphasizing the growing importance of sharing automated mobility concepts in the future (May et al., 2020). In this framework, SAVs and their associated business models offer promising solutions. In the use case of ride-pooling, SAVs could be a helpful complement in the middle of the two poles of public and private transport by combining similar routes of multiple individuals into one shared journey. The sharing of the trips and the absence of driving personnel suggest a low cost of travel. Although this is primarily a clear advantage of SAVs, at the same time it poses a latent danger to urban transportation systems, as low-priced alternatives to

In D. de Waard, V. Hagemann, L. Onnasch, A. Toffetti, D. Coelho, A. Botzer, M. de Angelis, K. Brookhuis, and S. Fairclough (2023). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2023 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

public transport could lead to a cannibalization of the latter (Krueger et al., 2016). It is therefore highly important to design SAVs carefully and to choose attributes, like interior features or overall comfort level, in such a way that primarily private car users are persuaded to use the service. This necessitates very precise knowledge of the needs and requirements of various user groups. According to Dorynek et al. (2022), it cannot yet be observed that the vehicles currently used to provide manually operated ride-pooling services meet these requirements and makeshift modifications by operators are the consequence. This suggests that as of now, insufficient attention is being paid to the vehicle interior and the aspect of user-oriented design is being neglected by manufacturers.

Limitations of SAV pilot experiments

In recent years, SAVs have been tested under real traffic conditions in numerous pilot projects. The experimental operation of these novel vehicles makes an important contribution to their faster implementation as potential users are able to experience automated rides first-hand and provide researchers with valuable insights for instance regarding their acceptance of the technology (Riener et al., 2019), experience on board (Dreßler & Höfer, 2022) or issues regarding safety and security (Salonen, 2018). However, this approach to user opinion acquisition also has its drawbacks. For example, Heikoop et al. (2020) report that the limited speed of the vehicles is perceived by the users as problematic. In addition, the prototypical setting of the trips with usually short route lengths as well as the mandatory safety operator on board only allow limited conclusions to be drawn about the later fully autonomous use of the vehicles over a longer trip duration. Kolb et al. (2020) even argue that the shuttle buses under consideration should be assigned to level 2 according to SAE (2018) due to the permanent monitoring required. Given these limitations, combined with the need for upscale comfort levels in automated mobility services, there is a strong demand for research that addresses SAVs of varying comfort levels in highly realistic scenarios. Even though experimental settings in virtual reality already provide an ever-increasing visualisation quality and thus represent a good experimental solution, the use of the WoOz technique is still the one alternative in this issue that is able to provide the most realistic experience for users. In the automotive context, this means that a participant interacts with what appears to be an automated vehicle. The seemingly automated system is however operated by one or more human agents who covertly operate the vehicle and its interaction with the passenger.

Previous research and current shortcomings of WoOz research in SAVs

The practice of WoOz experiments already has a long tradition in the field of automated vehicle research. From a methodological point of view, the considerations of Baltodano et al. (2015) provide an important groundwork for this paper. With the Real Road Autonomous Driving Simulator (RRADS), the authors introduce a framework for conducting inexpensive WoOz studies, using regular, commercially available passenger cars. They manage to convincingly hide the driving wizard behind a partition separating the driver and passenger seats, which largely eliminates the need for structural modifications to the vehicle and significantly reduces the cost of the experiment. In addition, they present a protocol procedure that serves as a guideline for the chronological sequence of a WoOz study.

In another significant study, Meurer et al. (2020) report about providing subjects with an on-demand robotaxi service for one week, operated as a WoOz trial. This is particularly noteworthy as the experimental setting is one of the very few WoOz studies conducted specifically on SAVs and because it is largely integrated into the test subjects' everyday life hence the rides they order are individual as well as actually necessary and useful. One important finding in their work is that participants wish to receive real time information about the ride in terms of upcoming stops and delays but also about the current traffic situation and potentially hazardous incidents. A similar approach is described in Alpers et al. (2020), who refrain from integrating the test drives into the daily routine of their participants, but instead apply a more elaborate interior design to the WoOz vehicle. They also report a clear advantage of a higher level of information – here in the form of a digital assistant.

Overall, only very few literature can be found that reports the application of the WoOz method in the context of SAVs. Previous research in this area focuses mostly on user experience than the methodological aspects that such experimental settings entail. At the same time, methodologically oriented papers regarding WoOz in automated driving exclusively involve passenger cars (Baltodano et al., 2015; Wang et al., 2017), which limits transferability of their findings to the area of shared mobility. Though some basic considerations certainly also apply to SAVs, there is a strong need for an alignment and an extension of the methodological framework specifically considering shared mobility in WoOz studies. The aim of this paper is to point out major methodological challenges in said experiments by conducting a WoOz pretest.

Method

Test vehicle and modifications

The seven-seater version of a Hyundai Staria was chosen as test vehicle. Its exterior appearance was modified in three ways. The word mark of a fictitious ride-pooling provider ("muvit") was attached as a sticker to the centre of each side of the vehicle. Two more stickers with the words "self driving shuttle" were placed on the sides of the last seat row to increase the credibility of the vehicle's autonomy. On the right sliding door, there was a sticker representing a camera lens, which served as a mock-up entry control. Test subjects had to hold a QR code in front of it to gain access to the shuttle. Tinted side windows at the front of the vehicle to prevent views of the driving wizard's workplace as in Alpers et al. (2020) were not feasible due to legal limitations in Germany.

The largest modification to the vehicle's interior was a partition made of lightweight foam boards between the driver's and the second seat row which ensured that the researchers could not be seen operating the vehicle or the interaction prototypes. In addition to two word mark stickers on the left and right of the partition, there were also two stickers representing emergency stop buttons. These were placed in such a way that they were easily accessible from the left and right seat respectively. For video recordings of the participants, an action camera was mounted at mid-height in the left corner between the partition and the wall of the vehicle. The test vehicle was equipped with a built-in camera that allowed observation but not recording of participants throughout the experiment. In the middle, right-hand centre of the partition, two tablets were mounted next to each other at approximately head height, which served as monitors for a wireless webcam broadcast of the current traffic situation (left) as well as to display information about the journey (right). To reinforce the impression of a ride-pooling service, magazines were placed in the map pockets for the participants to read during the ride. Figure 1 provides an overview of the vehicle's seat layout and the respective modifications made. Images of the vehicle's exterior as well as its interior can be taken from figure 2.

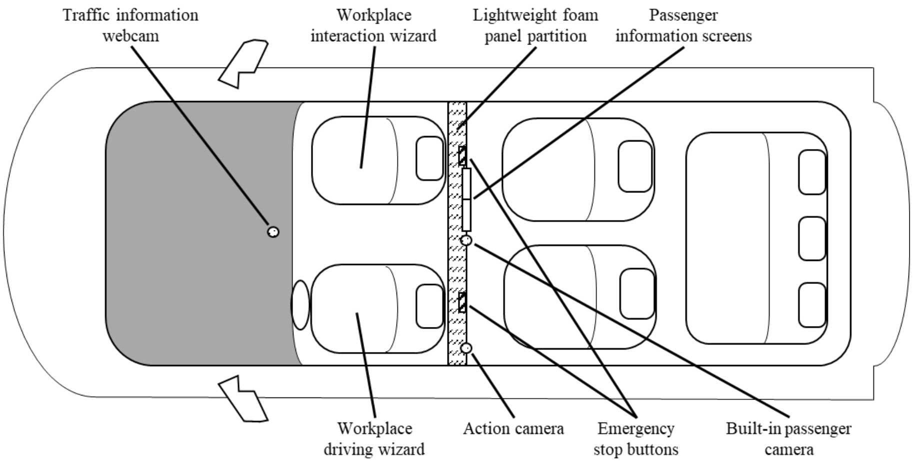

Figure 1. Seat layout and modifications of the test vehicle's interior.

Figure 2. The exterior (left) and the interior (right) of the WoOz test vehicle.

Interaction prototype

Participants' interaction with the mobility service in general and the vehicle in particular mainly involved two phases. These were (1) the booking process before entering the vehicle, when participants ordered a ride by using a prototypical mobile phone application and (2) during the ride, where two screens and a voice assistant provided information about the current traffic situation as well as the journey. The prototype for the mobile phone application was created with the UX prototyping software Figma and represented a mobility-as-a-service application that offered a choice of transport options including the "muvit" ride-pooling service. Participants were able to go through the entire booking process on a smartphone and finally book a ticket, whereupon they were presented with a QR code that served as identification when boarding the vehicle. During the ride, participants could observe a webcam-broadcast of the current traffic situation on the left of the two screens. In addition, as suggested by Fröhlich et al. (2019), a streaming software was used to show the vehicle's current speed and highlight relevant events or road users by placing overlays in real time. The right of the two screens had the function of welcoming passengers, displaying the route and current position of the vehicle, and providing general information about the journey. Figure 3 gives an impression of the different interfaces used in the experiment.

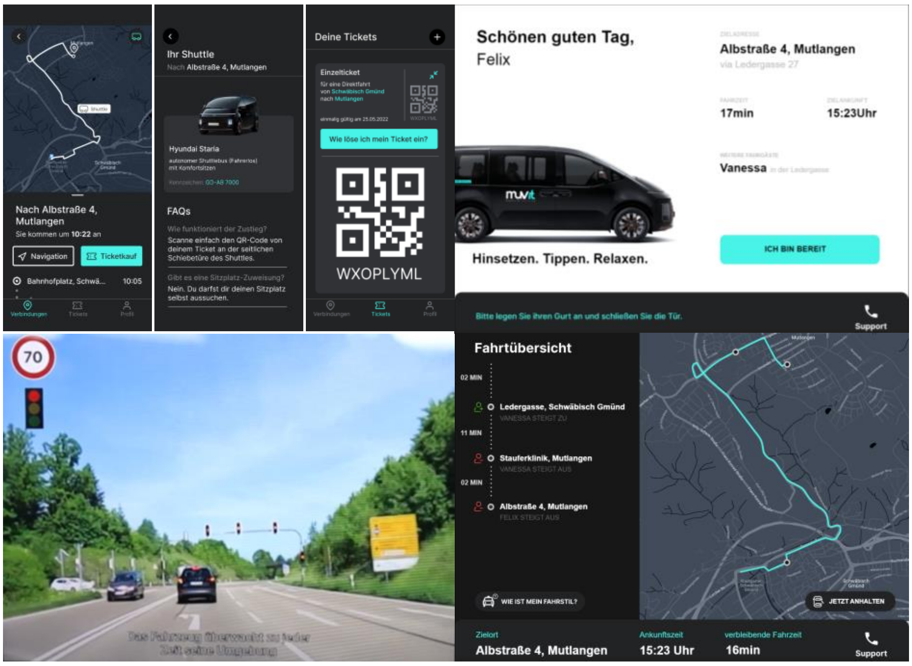

Figure 3. Mobile booking app (top left), passenger welcome screen (top right), traffic information screen (bottom left), and route information screen (bottom right).

Test route and study protocol

The experiment took place in May 2022 in the German city of Schwäbisch Gmünd with n = 8 participants. A driving route of approximately 15 minutes was chosen between the train station of Schwäbisch Gmünd and a suburban residential area in the neighbouring municipality of Mutlangen. As one of the aims of the trial was to integrate a multi-user component into WoOz testing, two additional stops, one shortly after the station and the other at a clinic in Mutlangen, were set along the route, at which a second participant got on and off. The route and relevant events during the journey are shown in figure 4.

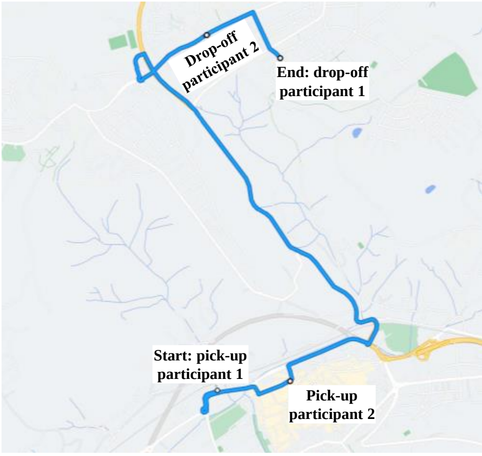

Figure 4. Route map of the experimental drive and relevant events.

The course of the trial can be divided into three phases: before, during and after the ride. Each run required two participants, who waited at separate locations for the trial to begin (participant 1: starting point; participant 2: pick-up point). Participant 1 was introduced to the testing procedure by the interaction wizard and was subsequently told to imagine they had just arrived at the station by train but missed their connecting bus. As an alternative to waiting for the next regular bus, participant 1 was instructed to book a ticket for a ride with an autonomous ride-pooling service on a smartphone they were given for this task. After the participant had booked the ticket, they approached the test vehicle from behind and entered it by presenting the QR-code to the camera lens on the sliding door. The interaction wizard took the front passenger seat, which was justified with legal requirements. A female voice assistant played over the vehicle's speakers welcomed the participant, instructed them to take their seat and put on a seatbelt. After safely stowing possible luggage, the participant tapped the "Start now" button on the welcome screen, whereupon the driving wizard began the journey. Throughout the entire ride, the interaction wizard triggered different voice outputs and integrated digital overlays into the traffic video stream. After a short ride of two minutes, the voice assistant announced that another passenger is going to join and after the vehicle stopped, participant 2 entered by presenting their QR code that had been sent to them earlier in order to minimize effort in the scope of this pretest. Once ready to start their ride, they also tapped the start button and the trip continued for approximately eight minutes until participant 2 reached their destination and exited the vehicle. After another minute of riding alone, participant 1 also reached their final stop and left the shuttle. In the post ride phase of the experiment, participant 1 was debriefed and offered a ride back to the starting point during which participant 2 was also picked up. Subsequently, both participants were asked about their experience with the ride. For this pretest setting, the opinions of the subjects were collected solely in the context of a short qualitative interview in which they were asked what should be continued, improved, stopped or added to the ride-pooling service.

Results

Qualitative post-ride interviews

Participants were positive about the comfort level of the service and the possibility to determine the departure of the vehicle by pressing the start button. The ingress process by presenting the QR-code was also mentioned as user friendly. Suggestions for improvement addressed the information screen display. A wish was expressed for the route progress information to be displayed more clearly during the journey. One point of criticism was that the live broadcast of the traffic ahead could be annoying or suggest that the passenger’s attention is necessary. In addition, it was not always clear to the participants where exactly they should get off. A clearer marking of the exit location in the app was mentioned as a solution to this problem. Additional features that participants wished for were more space between the two seats and a table for longer journeys. Stronger individualisation of the digital displays was also suggested. Lastly, it was noted that a physical emergency stop switch instead of a sticker could provide more realism. With regard to the credibility of the vehicle's autonomy, there were mixed results. While some of the participants were quite convinced that they were sitting in an actually automated vehicle, others were at least sceptical due to their background knowledge of the current state of the art of automated driving.

Lessons learned

Conducting a feasibility study regarding a WoOz trial in a multi-passenger setting provided important insights and highlighted critical requirements for such testings. It can be assumed that there are distinct differences between WoOz testing for private passenger cars and SAVs. Based on the experiences made, the following basic guidelines for WoOz testing of SAVs are suggested:

1. Vehicle choice: Ideally, a larger van or minibus is used to convey the feeling of a purpose built vehicle that does not create the impression of a private passenger car.
2. Vehicle interior and positioning of participants: A generous amount of space should be offered which entails the avoidance of bench seats unless the examination of lower comfort interiors is deliberately aimed at. A back seat setup as is already common in representing level 5 automation (Bengler et al., 2020) is advisable. A partition that separates the front row from the rear of the vehicle provides a comfortable workplace for the interaction as well as the driving wizard.
3. Consideration of the holistic travel process: In order to map the entire process of a journey with an autonomous ride-pooling service, the associated booking process should also be integrated into the experiment. This includes both the prototypical interface of a booking app and a frame story that gives participants a comprehensible reason for their journey. Similarly, the experiment should not end inside the vehicle, but the drop-off process including complete egress should also be part of the investigation.
4. Consideration of multi-user settings: SAVs will only be able to contribute to reducing traffic volume if they are in fact shared by multiple passengers.

Therefore, WoOz studies in the context of SAVs should be designed in such a way that realistic interaction between multiple passengers is possible.

Discussion

Based on a review of current literature, this pretest is the first of its kind explicitly aimed to explore difficulties in realizing a WoOz study in the context of SAVs and to identify the prerequisites for it as well as establishing guidelines by focussing on methodological aspects like vehicle interior modification, interface design, or interaction among passengers. Essentially, it can be stated that this experimental method is well suited for evaluating the needs of future users and it should find a larger place in the repertoire of user-centred shared mobility research. However, it must be taken into account that established procedures from the field of WoOz research on passenger cars should not just be blindly adopted. While methodological problems such as the consistency of the driving wizard's driving style are certainly also a challenge, other areas specific to SAVs must be taken closely into account.

First, the selection of the test vehicle plays a greater role. The vehicle interior needs to receive more attention in the development of purpose built vehicles (Dorynek et al., 2022). Therefore, by using spacious cars, research in future WoOz studies for SAVs could focus particularly on ingress-egress processes with multiple passengers, as well as on privacy or in-vehicle security issues. Also imaginable are studies on alternatives to the classic seating layout, such as campfire or vis-à-vis arrangements. Second, interaction concepts like passenger information screens or voice assistants as well as a frame story that justifies the trip are supposed to immerse participants as much as possible. As was shown by Baltodano et al. (2015) and Alpers et al. (2020) the use of a welcoming digital assistant improved participants' attitude towards the vehicle. In this context, the recommendations of Large et al. (2019) regarding conversational user interfaces should be followed. Finally, the ideal solution in terms of a naturalistic usage of a ride-pooling service would be the approach by Meurer et al. (2020) to integrate the experiment into participants' daily life. However, this is hardly feasible especially if more than one participant is supposed to take part in the experiment at the same time. Nevertheless, a multi-user setting is indispensable in the development and testing of SAVs. As long as this type of vehicle is not tested in the shared state under real road conditions, it is difficult to draw reliable conclusions about the behaviour and experience of future users in shared interiors. It is debatable whether the additional passenger should be another naïve participant or a confederate researcher. While it is true that the simultaneous study of two participants in the same vehicle presents a methodological challenge and greater susceptibility to the influence of confounding variables, this way of conducting WoOz studies would represent the most realistic way of conducting shared rides.

Bengler et al. (2020) rightfully call for strictly monitored consistency in WoOz studies, especially in the characteristics of the driving wizard. This certainly also applies to studies of this kind on SAVs. Additionally, to counteract the methodological issues of WoOz experiments, results from these studies should always be compared to and put in relation with findings derived from other research methods like virtual reality studies, interviews, questionnaires or workshops. Still, the outcomes of this pretest can be described as positive and increased usage of this method will certainly contribute to a more user-centred design of SAVs.

Acknowledgements

This work was funded by the German Federal Ministry for Digital and Transport as part of the research project SAVeNoW (FKZ: 45AVF2010M).

References

Alpers, B.S., Cornn, K., Feitzinger, L.E., Usman, K., So Yeon, P., Bardia, B., Hills-Bunnell, D.J., Trevor, H., Kaustubh, D., Rieko, Y., Larry, L., & Aquino Shluzas, L. (2020). Capturing Passenger Experience in a Ride-Sharing Autonomous Vehicle. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 83–93). New York. ACM.

Baltodano, S., Sibi, S., Martelaro, N., Gowda, N., & Ju, W. (2015). The RRADS platform. In G. Burnett, J. Gabbard, P. Green, S. Osswald, A. Eren, and V. Antrobus (Eds.), *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 281–288). New York. ACM.

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

Dorynek, M., Derle, L.-T., Fleischer, M., Thanos, A., Weinmann, P., Schreiber, M., Schumann, S., Tunc, T., & Bengler, K. (2022). Potential Analysis for a New Vehicle Class in the Use Case of Ride-Pooling: How New Model Developments Could Satisfy Customers and Mobility Makers. *Vehicles*, *4*, 199–218.

Dreßler, A., & Höfer, E. (2022). User experience of a self-driving minibus - reflecting vision, state and development needs of automated driving in public transport. In D. de Waard, S.H. Fairclough, K.A. Brookhuis, D. Manzey, L. Onnasch, A. Naumann, R. Wiczorek, F. Di Nocera, S. Röttger, and A. Toffetti (Eds.) (2022). *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2022 Annual Conference* (pp. 51-61). Downloaded from http://hfes-europe.org (ISSN 2333-4959)

Fröhlich, P., Schatz, R., Buchta, M., Schrammel, J., Suette, S., & Tscheligi, M. (2019). “What’s the Robo-Driver up to?” Requirements for Screen-based Awareness and Intent Communication in Autonomous Buses. *I-Com Journal of Interactive Media*, *18*, 151–165.

Heikoop, D.D., Nuñez Velasco, J.P., Boersma, R., Bjørnskau, T., & Hagenzieker, M.P. (2020). Automated bus systems in Europe: A systematic review of passenger experience and road user interaction. In *Advances in Transport Policy and Planning. Policy Implications of Autonomous Vehicles*, 5, pp. 51–71.

Kolb, J.C., Wech, L., Schwabe, M., Ruzok, C., & Trost, C. (2020). Technische Aspekte des automatisierten Fahrens am Projekt des autonomen Shuttlebusses in Bad Birnbach. In A. Riener, A. Appel, W. Dorner, T. Huber, J. C. Kolb, and H.

Wagner (Eds.), *Autonome Shuttlebusse im ÖPNV*, pp. 57–91. Springer Berlin Heidelberg.

Krueger, R., Rashidi, T.H., & Rose, J.M. (2016). Preferences for shared autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, *69*, 343–355.

Large, D.R., Burnett, G., & Clark, L. (2019). Lessons from Oz. In C.P. Janssen, S. F. Donker, L.L. Chuang, and W. Ju (Eds.), *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (pp. 335–340). New York. ACM.

May, A.D., Shepherd, S., Pfaffenbichler, P., & Emberger, G. (2020). The potential impacts of automated cars on urban transport: An exploratory analysis. *Transport Policy*, *98*, 127–138.

Meurer, J., Pakusch, C., Stevens, G., Randall, D., & Wulf, V. (2020). A Wizard of Oz Study on Passengers' Experiences of a Robo-Taxi Service in Real-Life Settings. In R. Wakkary, K. Andersen, W. Odom, A. Desjardins, & M.G. Petersen (Eds.), *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (pp. 1365–1377). New York. ACM.

Riener, A., Appel, A., Dorner, W., Huber, T., Kolb, J.C., & Wagner, H. (2019). *Autonome Shuttlebusse im ÖPNV: Analysen und Bewertungen zum Fallbeispiel Bad Birnbach aus technischer, gesellschaftlicher und planerischer Sicht*. Springer Berlin Heidelberg.

SAE (2018). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles* (SAE J3016). 400 Commonwealth Drive, Warrendale, PA, United States. SAE International.

Salonen, A.O. (2018). Passenger's subjective traffic safety, in-vehicle security and emergency management in the driverless shuttle bus in Finland. *Transport Policy*, *61*, 106–110.

Wang, P., Sibi, S., Mok, B., & Ju, W. (2017). Marionette: Enabling On-Road Wizard-of-Oz Autonomous Driving Studies. In B. Mutlu, M. Tscheligi, A. Weiss, and J.E. Young (Eds.), *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 234–243). New York. ACM.