A Wizard of Oz Field Study to Understand Non-Driving-Related Activities, Trust, and Acceptance of Automated Vehicles

Citation for published version (APA):

Document license:
Other

DOI:
10.1145/3409120.3410662

Document status and date:
Published: 21/09/2020

Document Version:
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
A Wizard of Oz Field Study to Understand Non-Driving-Related Activities, Trust, and Acceptance of Automated Vehicles

Henrik Detjen  
University of Duisburg-Essen  
Essen, Germany  
henrik.detjen@uni-due.de

Bastian Pfleging  
Eindhoven University of Technology  
Eindhoven, The Netherlands  
b.pfleging@tue.nl

Stefan Schneegass  
University of Duisburg-Essen  
Essen, Germany  
stefan.schneegass@uni-due.de

ABSTRACT
Understanding user needs and behavior in automated vehicles (AVs) while traveling is essential for future in-vehicle interface and service design. Since AVs are not yet market-ready, current knowledge about AV use and perception is based on observations in other transportation modes, interviews, or surveys about the hypothetical situation. In this paper, we close this gap by presenting real-world insights into the attitude towards highly automated driving and non-driving-related activities (NDRAs). Using a Wizard of Oz AV, we conducted a real-world driving study (N = 12) with six rides per participant during multiple days. We provide insights into the users’ perceptions and behavior. We found that (1) the users’ trust a human driver more than a system, (2) safety is the main acceptance factor, and (3) the most popular NDRAs were being idle and the use of the smartphone.

CCS CONCEPTS
- Human-centered computing → Human computer interaction (HCI); Field studies.

KEYWORDS

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

AutomotiveUI '20, September 21–22, 2020, Virtual Event, DC, USA  
© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-1-4503-8065-2/20/09 .. $15.00  
https://doi.org/10.1145/3409120.3410662
crucial to analyze and support the design of future vehicles from a user perspective.

Previous research on the behavior of passengers in automated vehicles (abbr.: AVs) is mostly limited to observations of other modes of transportation, e.g., trains or subways [35], or to potential users imagining such rides in interviews, surveys, or ideation sessions (e.g., [42]). Investigating the actual user behavior in such situations is still missing yet essential to understand how people spend their travel time in order to adapt user interfaces to the needs and activities of passengers [25]. This understanding is even more important as we know from the task-artifact cycle [8] that tasks / demands and artifacts co-evolve continuously.

1.1 Contribution Statement

In this paper, we contribute qualitative insights from a real-world driving study (N = 12) in a Wizard of Oz automated car with six rides per participant. In doing so, we bridge the gap to prior research, which investigated NDRAs during automated driving based on hypothetical scenarios and interviews. These findings help to better understand the future use of AVs and identify requirements for AV in-car user interfaces. Moreover, we report the users’ (change of) attitude regarding trust and technology acceptance in the real world.

2 INVESTIGATING NON-DRIVING-RELATED ACTIVITIES

Highly automated driving (HAD) is a complex phenomenon and investigating user interface aspects is especially challenging due to the chicken-and-egg problem[34, 49]: As automated vehicles are not yet available, users have difficulties imagining driving in such cars (e.g., when asked in surveys) and user needs cannot yet be extracted from real-world use cases. Therefore, researchers try to apply different methods and triangulate their results to get a comprehensive picture of the technology [43]. In the following, we provide an overview of research approaches to investigate NDRAs and show how a real-world driving experiment can influence the users’ judgment regarding AVs.

2.1 (Non-)Driving-Related Activities

Bubb [6] distinguishes three types of driver tasks in a conventional car. The primary driving task comprises all required actions to reach a particular destination. Secondary tasks support the primary task but are not necessary to reach the destination (e.g., use the turn indicator). Tertiary tasks are not related to transportation (e.g., infotainment control). With increasing automation, the role of the human in the car shifts from being the driver to being the passenger. In this situation, the former primary and secondary driving tasks fade away, swapping importance with the former tertiary tasks. To avoid confusion when using these established terms, we follow the terminology of Pfleging & Schmidt [35] who propose to use the terms of Driving-Related Activities or Tasks, which include the former primary and secondary tasks, and Non-Driving-Related Activities (NDRAs), which include the former tertiary tasks.

2.2 Methods to Investigate NDRAs

Up to now, prospective users of AVs have no or only little real-world experience with driving automation. Hence, researchers need to create a vivid mental image in the users’ minds of such a not yet available technology when investigating user aspects in hypothetical use cases. The higher the users’ immersion for the technology scenario is, the better we expect to obtain future needs and requirements. Looking at existing research, we identified four major methodological approaches on how to study NDRAs.

While the remaining categories of approaches are based on observation of the passengers’ behavior, the first category is based on self-reports after thought experiments [41]. These studies commonly use a scenario description of HAD and ask the participants about the hypothetical situation, e.g., in design sessions [42], interviews [31], or online questionnaires [10, 34, 40]. They depend on the understanding and sensitivity of the participants. A definite challenge for such approaches is the Chicken-and-Egg Problem [49] of HAD: As automated cars are not yet publicly available, there is no public experience with such cars. This lack of experience makes it difficult for participants to imagine a future (hypothetical) usage scenario. Referring to the task-artifact cycle [7], one typically builds technology to fulfill existing human needs and preferences (for AVs, this could refer to the needs for autonomy, security, stimulation, and meaning [14]). The availability of technology influences and changes the users’ needs and preferences, which leads to a continuous cycle. Consequently, extracted findings from thought experiments may serve as initial input for the development of AVs. However, these may only represent a very first snapshot of user needs and preferences.

The second category of approaches centers around studies of driving simulator experience in classical driving simulators [17]. In the future, also video-based driving simulators [15] seem feasible to investigate NDRAs. While such experiments offer a safe environment, an essential drawback of such approaches is that they are conducted in an artificial lab environment and that (at least static) driving simulators may create a different perception of driving (e.g., related to motion sickness). We expect both to influence the users’ choices for NDRAs.

The third categories is closer to reality and comprises studies based on the observation of experiences in other contexts, which includes other modes of transportation such as trains [34, 38], busses [38], or taxis [32]. While findings of such experiments can be a starting point for investigating NDRAs, an essential drawback of investigating user behavior and needs for NDRAs in such environments is a lower level of privacy. We expect this to have a definite impact on passenger activities as users might avoid certain activities in more public environments or choose different activities.

The final category comprises studies that base on real-world or in-situ experience. Real-world experiments using actual AVs may lead to severe risks for drivers, passengers, and other road users [5], which needs to be avoided. One can increase safety in such experiments by regulating and controlling the system or the environment. One can imagine to a) use a real AV in a simulated environment, or b) use a simulated AV in a real environment, or c) use a mix of both. Through virtual reality (abbr.: VR) driving simulators in real
cars [16, 19] a fully controlled environment would be possible. However, they are not feasible to study NDRAs because they require the user to wear a head-mounted display that impedes real-world physical activities. A virtual and controlled environment can also be created through windscreen augmentations [3] or video projection walls [26]. Here, the passenger is free to do real-world tasks in the car, yet the user experience might be limited, e.g., driving speed or traffic. Wizard of Oz Vehicles [2, 22, 28, 47, 48], provide a realistic driving environment in a real car. Such approaches include the deception of the users by giving them an illusion to interact with a real system (i.e., drive in a fully automated car) while in fact the system is operated by a human operator (“wizard”). If the illusion is maintained, this method allows a rich user experience concerning the use of future AVs.

Building upon the task-artifact cycle [8], our goal is to go beyond prior work, which mainly investigated the first set of user needs and preferences before the existence of an actual product. Thus, we want to understand user needs, preferences, and behavior based on using an actual vehicle that resembles an automated car’s features. Therefore, we decided to create an in-situ experience through a WoZ automated car to study NDRAs, accompanied by interviews and self-report questionnaires to study further phenomena such as automation trust and acceptance. By repeatedly exposing the participants to this vehicle through multiple rides, we expect the user to acclimatize themselves to the vehicle and, thus, activities and attitudes to change over time. By conducting a study in real life, we are much closer to investigating user behavior in AVs than it was possible with other studies as the ones mentioned above.

2.3 The Effects of a Real-World Driving Experience on User Attitude

Looking at the related work, many models investigate and explain automation acceptance or trust exist. A common drawback of most studies is again that they are based on thought experiments. As we expose participants to an actual car in real-world scenarios in our experiment, we also take this as an opportunity to validate results from existing models with real-world data from our experiment.

2.3.1 Acceptance of Automated Vehicles. Popular acceptance models like the Technology Acceptance Model [45] or the Unified Theory of Acceptance and Use of Technology [46] use factors like effort expectancy, performance expectancy, social influence, and facilitating conditions to predict the intention to use a system. Adapted for the automotive context, the Car Technology Acceptance Model [29] and the Automated Vehicle Acceptance Model [18] add domain-specific factors like anxiety, perceived safety, and attitude towards using the technology. Rödel et al. [37] also use the user experience factors trust and fun. Payre, Cestac & Delhomme [31] complement interest in impaired driving as an essential factor. Most factors come from interviews or surveys based on text-scenarios. We expect to collect new insights by inquiring users who experience real-world AV. The same applies to the measurement of trust in automation.

2.3.2 Automation Trust. Trust is an essential factor when it comes to the adoption of AVs [9]. Although, people’s trust lowers with higher automation levels [37]. Consequently, we expect to get low trust ratings within our study for the experience with our WoZ vehicle. Trust is a dynamic construct [27] and can be divided into three layers [20]: 1) dispositional trust based on personality and socio-cultural aspects, 2) situational trust based on the interaction context, and 3) learned trust based on experience with a system over time. In our study, we do not influence dispositional or situational trust, but we expect an increment of learned trust over time.

2.4 Summary

Summarizing related work, we see a lack of experiences from real-world encounters between users and automated vehicles involving preferences and usage patterns for NDRAs in AVs. Besides, due to the non-availability of actual AVs, we still see a research gap concerning understanding aspects of acceptance of AVs and the users’ trust in the automation when being repeatedly exposed to such cars in real life. With our real-world experiment, which uses a Wizard of Oz automated car, we want to bridge this gap.

3 REAL WORLD STUDY IN A WIZARD OF OZ VEHICLE

The core part of our study is a real-world observation of users in a WoZ vehicle throughout six subsequent rides per participant over multiple days. These rides are accompanied by semi-structured interviews and questionnaires. In addition to adhering to current WoZ reporting guidelines [36], we implemented the advisory recommendations of the departmental ethics committee in Eindhoven.

3.1 Participants

We recruited 12 participants (6 female, 6 male) aged between 24 and 33 years ($M = 28.67$, $SD = 3.08$) from personal networks, via local mailing lists, and personal address. We selected participants with different backgrounds (see the list below, participants marked with * noticed the deception) in order to get a broader range of impressions:

(1) Emma (f, 20-25, Media Designer)
(2) Elias (m, 31-35, Freelance Programmer)
(3) Isabella (f, 20-25, Waitress / Student)
(4) Sophia* (f, 26-30, Barkeeper)
(5) Charlotte (f, 26-30, Job-Seeking)
(6) James (m, 26-30, Student)
(7) Lucas* (m, 31-35, Freelance Musician / Student)
(8) Mia (f, 26-30, Teacher)
(9) Alexander* (m, 31-35, Service Engineer)
(10) Jacob* (m, 26-30, Research Assistant)
(11) Olivia (f, 26-30, Financial Consultant)
(12) Benjamin (m, 20-25, Material Tester)

The participants’ affinity for technology, measured with the ATI scale [13] (9 items on a 6-point Likert scale), was rather high ($M = 3.87$, $SD = 1.24$, $MIN = 2.00$, $MAX = 5.67$). Participants had a regular daily commuting/travel time of short to medium-length ($M = 19.63$ min, $SD = 10.49$ min, $MIN = 5.6$ min, $MAX = 50.28$ min) which we expect to be most common for the use of AVs in everyday life in Europe (e.g., trip to work or to the next shopping center). We conducted the study in the urban areas of Bremen and Essen in Germany. As the participants received the six
commutes/rides during the experiment free of charge, they did not receive additional financial compensation.

3.2 Wizard of Oz Vehicle

We used the TU/e Mobility Lab [24], a Renault Espace van, and modified it in a similar way as the RRADS platform [2] to create the illusion of an AV. We refer to the actual driver of the vehicle as “automation wizard” or “safety driver”. A separating wall hides the safety driver form the participant who is sitting on one of the rear seats (see Figure 2c to Figure 2d). Thus, the wall separates the safety driver space from the passenger space.

The separating wall carries a 43” 4K display (Sony KD43XF7596) that connects to a GoPro Hero 6 Black action camera mounted at the inner top center of the windshield. The display provides the passengers with a 4K real-time view of the driving scene as if they were sitting in the (co-)driver seat. To increase passenger comfort and context awareness, vertical LED bars behind the left and right edge of the screen indicate if the car is turning left or right. The passenger space comprises standard car seats, a wooden center armrest, and a central table, which the passengers can use, for example, to place devices such as laptops and tablets or other personal belongings and catering (see Figure 2b). Participants were free to bring their own devices and connect them with an in-car WiFi hotspot.

Participants could use an “emergency” button on the central table to signal discomfort and the need for action (e.g., to stop the car). In order to monitor and record the passengers’ behavior, we used a laptop on the co-driver seat, which connects to an additional camera above the TV screen on the cabin partition (see Figure 2d).

3.2.1 Automation Wizard. One of the authors of this paper served as the automation wizard (m, 32 years) and had full knowledge of the study objectives. To get accustomed to the car and develop an AV-like driving style, he practiced driving the car for one week before the start of the experiment. To minimize potential risks during the rides and to maintain the illusion of an automated car, the wizard driver imitated the expected defensive (e.g., not exceeding speed limits and being attentive to other road users) and forward-thinking driving style of an AV.

3.2.2 Cover Story. Since telling participants that we will observe them during the experiment might influence their behavior, we deceived the participants by telling a cover story. This deception also included that we pretended that the car is driving automated in order to be able to measure trust in automation and acceptance of automation technology.

We briefed the participants that the goal of the experiment is to investigate the influence of different driving styles of AVs on its users, regarding comfort and personal preferences. We instructed the participants to judge the randomly assigned driving style after each ride. By telling this, we covered potential inconsistencies in the wizard’s driving behavior. We explained that we would use the video recordings of the interior to understand the passenger’s reactions to the current driving behavior better and also use this real-time video feed as a tool for the safety driver to react to requests from the
participant. As the participants could see the automation wizard – also steering the car – before pick-up or after drop-off, we explained that the safety driver would a) continuously monitor the vehicle behavior and b) assist the automation if necessary and mainly for parking the car, i.e., around pick-up and drop-off. After the study, we asked the participants if they noticed the deception. Four persons (33\%) stated that they had noticed the deception (denoted by * in the participant list). For all other persons \((n = 8)\), the deception was maintained for the whole study.

### 3.3 Procedure

For our study, we followed the procedure as outlined in Figure 2a. During sign-up, we informed the participants that the study consists of six freely selectable rides with our AV along with introductory (kick-off) and final interview sessions. The participants received a flyer with information on how to book the six individual car rides.

3.3.1 Introductory Session and Deception. For the kick-off session (25 min), we visited the participants at their home or invited them for a video call, depending on their preference. Before they decided on their participation, we encouraged the participants to ask any questions related to their participation or the purpose of the project (cover story). We ensured that they received understandable and satisfactory answers. Complying with GDPR [12] guidelines, we asked for consent to use the data, also for follow-up research. Being a passenger in the WoZ car may lead to motion sickness, for instance, when reading or using mobile devices. However, we did not expect this risk to be higher than being a passenger in a standard car. Nevertheless, to prevent discomfort, we informed the participants that they could always withdraw their consent and ask the driver to stop the car at any time. Once we briefed the participants, we asked them to sign the consent form and handed out a copy to them. Next, they filled out the initial questionnaire containing questions about demographics (age, gender, education, job), mobility patterns, technology acceptance (including a scenario description and a picture of the WoZ car; measured with AVAM questionnaire [18], 26 items on a 7-point Likert scale), the affinity for technology and expectations about HAD (activities, inventory). Additionally, we conducted semi-structured interviews to capture attitudes towards HAD, e.g., related to acceptance and expectations.

3.3.2 Real-World Driving Study. The participants signed up for their six individual rides by providing the pick-up date and time, start, destination, and expected ride duration for each of their rides to the experimenter. At the indicated pick-up time, we picked up the participants at the start location and drove them to their destination. During the rides, we videotaped the passengers’ activities in the car. At the end of each ride, we asked the participants to fill out a short questionnaire about their automation trust (measured with the Trust Scale questionnaire [23], 12 items on a 7-point Likert scale), and, to make our cover story more credible, about their ride (style, comfort, safety, overall satisfaction).

3.3.3 Final Session, Explanation of Deception, and Debriefing. After the sixth and final ride, the participants filled out the post-questionnaire in the car. As a next step, we conducted a second semi-structured interview to detect changes in their attitudes in comparison to the initial interview. At the end of the interview, we debriefed the participants, explained to them the deception aspect of the study, and informed them about the real purpose of the experiment. We gave each participant the choice to maintain or withdraw their consent and the use of their data in case they do not agree with the actual goals and the purpose of the study, i.e., observing behavior and measuring trust/acceptance. None of the participants withdrew their consent.

### 4 RESULTS

The results of our study are threefold. First, in a thematic analysis [4], we collected insights into users’ expectations, how they perceive the impact of automated driving, and reasons for using or rejecting the technology. Second, we present the user’s perception of the AV in the real world, where especially trust in the vehicle capabilities, vehicle safety, and the individual use of time play an important role. Third, we report on our observations about the NDRAs in the WoZ automated car. As four participants (33\%) stated that they had noticed the WoZ deception we excluded their responses from the post-study interviews, trust and acceptance questionnaires, as well as from the activity observation to include only data with “real” automation experiences.

#### 4.1 Expected Impact of Automated Driving

From our interviews, we collected a multitude of impressions regarding the expected influence of HAD on society, groups, and individuals.

4.1.1 A better Life? Regarding the introduction of AVs in society, one participant uttered worries about the unknown situation (Olia: “Of course you have some respect for that, you just don’t know that someone else is steering for you.”) and think that life might be more digitized. On the other hand, participants also expect to have more time due to the removal of the driving task \((n = 2\), Mia: “Especially if you live in such a performance-oriented society, you can use the time to work more efficiently”) and better traffic regulation \((n = 4\), Benjamin: “There would hopefully be fewer traffic jams”). They perceive this extra time as comfortable \((n = 2)\), and expect that people will be less stressed \((n = 1)\). Traffic performance also influences the environmental effects \((n = 3)\) of mobility, which might lead to more energy efficiency, less CO2 emissions, and noise. Further, safety could be improved \((n = 1, Lucas: “There will be far fewer accidents”)\). Overall, a few participants thought that HAD would not affect their lives \((n = 3)\), but most expected their lives to be better \((n = 9)\).

4.1.2 Personal Autonomy and Free Time Through Automation. On a personal level, a few participants \((n = 2, Benjamin: “I’ll have to try it first”) are undecided, but most \((n = 10)\) think that AVs will offer more individual freedom (Sophia: “That you can do something else in time, that’s already a plus of freedom”). When asked what they would do with this new freedom, participants name especially the following activities: working \((n = 4)\), reading \((n = 4)\), eating and drinking \((n = 4)\), and doing nothing or relaxing \((n = 4)\). While one person explicitly names sleeping as an activity, another participant could not imagine sleeping in the car (Mia: “Maybe that’s still a little scary somehow because I’m not quite in control at all”).
4.1.3 The Car as a Social Place. Regarding future ride sharing or ride pooling, we asked participants how the presence of people known to them would influence their activities in an AV. Most participants \((n = 9)\) think that they will have more conversations (Sophia: “Because everyone can chat and the driver does not have to be excluded anymore”) or that they engage in other forms of social interaction like playing a game \((n = 1)\). Another aspect mentioned is the improvement of social events \((n = 2)\) if people are unfit for driving or if the event location it is not well connected to public transport (Jacob: “I think the celebration culture is changing, you can drive home after some drinks. From everywhere. That’s awesome”). Sometimes, the latter excludes persons who live in an area outside of the public transportation network. Furthermore, one participant mentions that sleeping might be an option in the presence of others because responsibility is shared (Mia: “I’ll be a little calmer”). If the other persons in the car are not known, people expect rides to be like if they were alone \((n = 2)\) on the train or bus \((n = 2)\). Sleeping in the presence of strangers is perceived as unsafe \((n = 2, Mia: “You have your valuables with you”).

4.2 Acceptance of Autonomous Vehicles

We analyzed and clustered responses from the interviews to identify essential factors regarding well-being and comfort in the car, as well as the main factors for the acceptance or rejection of AVs.

4.2.1 Well-Being and Comfort. The most frequently mentioned reason for well-being and comfort is safety – the feeling that a) the car acts safely \((n = 4, Benjamin: “A safe and controlled journey”) and b) it is made of robust technology \((n = 2, Charlotte: “Flawless technology”). Technology also needs to be understandable \((n = 1, Emma: “If you know how this works, then maybe I could relax too”). Comfort is related to the driving style \((n = 2)\). The defensive driving style in our experiment was perceived as comfortable (Benjamin: “Because it started relaxed and drove at a moderate speed, I found that very pleasant”). Some participants also name the vehicle interior \((n = 3)\) like a music system, Internet access, or massage seats, and vehicle characteristics like a silent engine \((n = 1)\) as important aspects for comfort. Overall, the gradually growing experience with an AV is expected to increase well-being and comfort \((n = 3, Mia: “I’m a little skeptical. That might settle over time”). After the experiment, Mia confirms this (“I have become [...] a little more open-minded and less skeptical”).

4.2.2 Key Factors and Barriers. Similar to comfort, when asked about the most important reason to use AVs, adequate safety is mentioned \((n = 6, Elias: “The feeling that the car is driving safely”). Furthermore, the gain of available time \((n = 1)\) and the relief from the driving task \((n = 1, Charlotte: “I don’t have to focus on traffic all the time”) are named as key factors. While participants consider not having to drive as a relief, some participants still prefer to keep some control of the car (Mia: “If I still have some way to intervene, I’d use it”). Furthermore, environmental reasons \((n = 1)\) and the ability to be mobile in a non-driving fit condition \((n = 1)\) are important use intentions. In the end, AVs should comparably be priced to conventional vehicles \((n = 2, Isabella: “That it is not disproportionately expensive”).

On the other hand, the participants identify some reasons to reject automated cars – sometimes being just the opposite of the acceptance reasons: If the car is perceived as unsafe \((n = 4)\), e.g., through media (Isabella: “If there had been news about accidents”) or when it is too expensive \((n = 1)\), some people’s willingness to use is lower. Additionally, if users cannot control the car or the driving style some expect a loss of autonomy (Mia: “If I now sit in front of the steering wheel and cannot intervene then I would probably not use it”) or a loss of driving pleasure (Benjamin: “It is still a matter of one’s own whether one can accelerate, brake oneself, drive around curves – or if the car does it for one”). Also, ethical considerations \((n = 1, Mia: “value conflicts”) of the vehicle behavior can influence the willingness to use.

4.2.3 Acceptance Before and After the Study. Figure 3a shows the acceptance factors measured with the AVAM questionnaire before and after the study. In general, participants rate performance expectancy, attitude towards automation, self-efficacy, and facilitating conditions rather high. Social influence and perceived safety are rated neutrally and anxiety rather low. Consequently, the behavioral intention to use AVs is rather high in our sample. Utilizing a dependent t-Test, we did not find any statistically significant differences between the scores before and after the experiment.

4.3 Trust in Automation

When asked about whom to trust more as a driver, human or machine, we see split opinions. While some people rather trust a human chauffeur \((n = 7, Elias: “An ultra-complex system always has a weakness”), others prefer an artificial autopilot \((n = 4, Charlotte: “Because I hope that human mistakes do not happen, e.g., when being tired”). Only one participant was undecided (Isabella: “Depends on the driver, I also feel uncomfortable in some lifts [...] and if they are super programmed, I would probably rather trust the machine than any person”).

The user’s trust in automation is expected to increase with experience \((cf., [20], Sophia: “Human. But only because I’m used to it”). To capture the experience, participants rated their level of trust after each ride. The measured trust rating is rather high on average \((M = 5.41, SD = .87)\) and slightly increases after the sixth ride (see Figure 3b). Further, we found strong correlations between the average trust rating of the rides and the social influence \((r(8) = .75, p < .05)\) and the intention to use the vehicle \((r(8) = .71, p < .05)\) from the post-questionnaire.

Also, we asked the participants whether they rated the safety driver, or the pure automation, or both in the questionnaires. As a result, we can further split the remaining group of eight participants who did not notice the deception in a subgroup that rated the pure automation \((n = 4, short: A)\) and a subgroup that rated a combination consisting of the automation and the presence of the safety driver \((n = 4, short: S)\). The S-subgroup \((M = 5.97, SD = .57)\) has higher trust ratings than the A-subgroup \((M = 4.85, SD = .79)\), however, these differences are not significant \((t(6) = 2.29, p = .06)\).

4.4 Passenger Behavior

To find the usage patterns of passengers in an automated vehicle, we analyzed the video recordings of the participants’ activities in
4.4 Passenger Behavior

To find the usage patterns of passengers in an automated vehicle, we analyzed the video recordings of the participants’ activities in the WoZ car and separately asked them what activities they expect to do when (use cases).

4.4.1 Activities During an Automated Ride. In addition to our observations, we asked the participants before and after the experiment to name five activities that they most likely would perform in an AV. The ranked order of their chosen activities is:

1. Using the smartphone (before: n = 9; after: n = 9)
2. Eating & drinking (before: n = 7; after: n = 6)
3. Reading (before: n = 7 after: n = 5)
4. Watching out of the window (before/after: n = 4)
5. Preparing for the job (before/after: n = 3)

Similarly, we asked for up to five items they would place in an AV and did so before and after the experiment. The five most frequently mentioned items are:

1. Cosmetics
2. Talk to Passenger
3. Eating and Drinking
4. Smartphone Typing
5. Writing

Figure 3: Results of the real-world study: Acceptance, trust, and observed NDRAs.
(1) Laptop / PC / tablet / gaming console (before: \( n = 6 \); after: \( n = 6 \))
(2) Pillow or blanket (before: \( n = 6 \); after: \( n = 3 \))
(3) Foldable table (before: \( n = 4 \))
(4) TV / computer screen (before: \( n = 4 \); after: \( n = 2 \))
(5) USB / power charging station (before: \( n = 1 \); after: \( n = 5 \))

The desired, predominant technical inventory corresponds to the observed activities.

In our video observations during the rides, we collected more than 23 hours (1424 min) of video material in total. For the categorization of observed activities, we use the proposed categories by Pfleging et al. [34] as a basis to simplify comparisons. We added new activities that we observed. To simplify the evaluation, we did not introduce a new category for parallel tasks, e.g., listen to music while chatting, but counted both activities separately. Furthermore, we ignored distractions from the main activity, which last under three seconds (e.g., short glance out of the window before switching back to the smartphone). Figure 3c shows the distribution of the activity category shares for all passengers who believed in the deception (\( n = 8 \)) concerning the total time. The top five activity found were (1) Watching out of the Window (55.47 %), (2) Smartphone Use (10.24 %), (3) Listen to Music / Radio / Audiobooks (9.05 %), (4) Office Tasks (8.17 %), and (5) Smartphone Typing (6.72 %).

4.4.2 Use Cases for Automated Rides. In general, people see benefits in using AVs for longer rides (\( n = 5 \)) such as cross-country trips (\( n = 1 \)), on the motorway (\( n = 1 \)), or more specific for holiday trips (\( n = 2 \)) like going to the sea with friends (\( n = 1 \)). Moreover, they would use it in situations where high concentration is required, e.g., in traffic jams (\( n = 2 \)) and for routine trips like the daily commute to work (\( n = 6 \), Benjamin after the experiment: “I would take advantage of it, just driving to work in the morning was much more pleasant than driving myself. I arrive at work much more relaxed”), to go shopping (\( n = 2 \)), or transport things (\( n = 1 \)). Another use case could be and is already partly the use in controlled and restricted environments (\( n = 1 \), Benjamin: “such as on a factory site”). Identified novel use cases are a) the transportation of children through AVs (\( n = 2 \), Mia: “I’d find it handy if you had kids if they could drive in there alone”) and b) the vehicle use under short-term limitations like sickness or intoxication (\( n = 3 \), Jacob: “When you want to come home from partying, but are somewhere where you cannot easily go home”) or under long-term limitations (\( n = 1 \), Sophia: “Physically or mentally limited people, that they can use the car and do not endanger traffic”).

Short trips (\( n = 4 \)) and city scenarios (\( n = 1 \)) are named as anti-use cases by some people (Isabella: “For such short things as “I just drive to the bakery” it would be too exhausting for me always to enter the destination because I am faster when I drive myself”).

4.4.3 Influence of the Safety Driver. After the experiment, we asked the participants if and how the presence of the safety driver influenced their behavior. The primary effect is that some people (\( n = 4 \)) were relaxed (Emma: “It felt more like a taxi ride, if you would sit in the front and the steering wheel and the pedals move, then that would be something else again”). Furthermore, the experimental situation was not always perceived as private and, thus, not suited for private communication, as stated by one user (Jacob: “I would have communicated more with my buddies if I had known that no one was listening to me”).

5 DISCUSSION
Our results reveal how users perceive AVs and how they may spend their in-car time. In the following, we discuss our findings, relate them to prior work, and indicate future research opportunities.

5.1 NDRAs - Intentions vs. Observations
In our study, we observe differences compared to the activities people mention in online surveys or design studies. Direct comparisons with activities from Pfleging, Rang, & Broy’s online survey [34] indicate similarities and differences to our observations. Figure 4 shows a direct comparison between activity expectations from their survey and the observed activity frequencies from our experiment.

We found the following similarities: The activity Watch out of the Window was ranked 3rd in their survey, with 82 % of the people expecting to do it (very) frequently. In our study, Watch out of the Window happened most of the time (55.5 %) and was done in 95 % of all rides. (The fear of) motion sickness could be one reason why the participants in our study dedicated a huge amount of their time just watching out of the window. However, this is subject to future research as we did not investigate the effect of motion sickness. Both the survey by Pfleging et al. and our observation found the usage of smartphones as a common activity. In their survey, the participants ranked Internet (61 %) and Social Media (48 %) rather high. Given that we did not observe the participants’ smartphone screens, we combined these categories. Then, we compared the survey category with our Smartphone Use category, which consumed 10.2 % of the time and was done in 67 % of all rides, which was the second-leading activity. Nevertheless, since this is one of the expected main activities, a more fine-grained analysis of smartphone use is necessary [30] and should be investigated in future work.

We identified the following differences compared to Pfleging, Rang & Broy: The top expected activity in their survey was Music, Radio, etc. (88 %). In our study, the corresponding category Listen to Music/Radio/Audiobooks was ranked third in total time share, but was observed in only 6.25 % of all rides. Similar differences are found for Texting, Eating & Drinking, Calling, Reading a Book, Sleeping, etc. There are two possible explanations for the differences. The first is the much smaller sample size in our study. The second is that the trip duration in our experiment was about 20 min on average. We expect this duration to be the average distance from home to work or to a friend living in the next city. Future work should take these influences on the reason for a trip taken into account. We argue that for short to medium distance trips, certain activities, like watching a longer movie or reading a book, are not likely to be performed in AVs and people usually think of long-distance trips when they think of automated driving. In general, existing descriptions of expected passenger activities are probably linked to long-distance trips, and future research should reflect the effect of trip duration / length by developing distance-related use-case surveys.

5.2 Acceptance Models for Automated Driving
We did not see changes in acceptance factors before (in form of a thought experiment experience through text/picture description of
the vehicle) and after the experiment (from participants’ real-world experience). An explanation could be that experience is not that important for the intention to use AVs, at least to some extent. For example, a study found that Tesla drivers kept using their system as usual despite being aware of system weaknesses [11]. Also, the presence of the safety driver might have an impact on these ratings, especially on perceived safety and anxiety. Existing car acceptance models like the AVAM [18] help to capture the public acceptance of AVs. However, to support the success of this technology, we believe that there must be a more explicit focus on the additional benefits and user needs that AVs can provide or support. For example, impaired driving, the use of the vehicle while being physically or mentally limited somehow, was found to be of high importance for some people and named as an important use case. One person drank a beer in our car on his way home after work. These observations are in line with findings of Payre, Cestac, & Delhomme [31], who found interest in impaired driving as a fundamental acceptance factor.

Such new needs should be included on top of acceptance models, which currently (still mainly) try to transfer the use case “driving and limited side-tasks” with corresponding needs on automated driving. However, as automated driving is expected to enable entirely new use cases like mobile office or sleeping, these possibilities need to be taken into account as these use cases produce new demands, which are entirely different from what was possible with manual driving. Looking at the observed activities and subjective feedback from interviews and questionnaires, we see one challenge in addressing different contexts like the trip duration (good predictor for suitability of use cases [1, 33], e.g., watching a movie requires a certain trip duration) or the user’s intention itself (e.g., entertainment, work, and flexibility).

Figure 4: Automated driving – Comparison of activities’ expected frequency (very frequently + frequently) found in the survey from Pfleging, Rang, and Broy [34] with frequency of activities found in this study (observed in n rides / all rides)
5.3 Experience and Trust

The results indicate that trust is higher when relying on the system with safety driver compared relying on the system alone. As a result of this, we can confirm previous research in a real-world setting, e.g., by Rödel et al. [37], who found through an online survey that user experience and trust decrease with higher levels of automation.

Moreover, the study suggests that the level of trust in the vehicle’s capabilities also influences the expected activities. Sleeping is mentioned to require a certain level of trust (also in other potential passengers). Hence, sleeping might be a valuable indicator to measure trust in automation because we expect people to only fall asleep in a (perceived) safe and controlled environment. On the other hand, this also indicates that activities might change over time when the level of trust increases.

5.4 Limitations

When investigating user aspects of future automated vehicles, researchers especially face challenges regarding the naturalness or technical fidelity of their studies on either the car side or the environment side. In our study, we prioritized validity on the environment side and, as a trade-off, chose to use a WoZ approach. An inherent weakness of our specific implementation is that the automation wizard can be uncovered. If this happens, the perception and behavior of the corresponding participants might be influenced and their data might need to be excluded as we did in our study. In addition, user behavior in a real AV might still differ due to different driving behavior.

Given the relatively small and young sample of our study, we see the observed activities as a first approximation of future behavior. Larger-scale follow-up research should verify how users perceive AVs and perform NDRAs in AVs. However, our findings provide valuable insights to inform these future experiments and concepts.

From a methodological perspective, we see that using a WoZ vehicle is a valuable and viable method to study qualitative aspects of HAD.

6 CONCLUSION

In this paper, we report on a field study (N = 12) of six short to medium duration rides that investigates user needs and expectations in a real-world setting using a WoZ automated vehicle. With this, we extend prior research on understanding user needs for automated driving, which so far mostly focused on observing people in other contexts (public transport) or interviewing potential users who had to imagine the automation situation. With our real-world experiment, we provide in-depth insights into AV usage behavior. In particular, we found that: (1) the participants indicated in their interview responses a higher trust rating for a human driver than for a system; (2) the acceptance did not change a lot over time and safety is named as the main acceptance factors by participants; (3) the most popular activities are being idle and the use of the smartphone while other activities occurred much less frequently than expected.

In summary, our results show that there is a difference between investigating user needs and preferences from surveys and real usage. With our field study, we therefore improve understanding the design requirements for future in-car interfaces in the context of NDRAs.

REFERENCES


