Building Collaborative Test Practices: Design
Ethnography and WOz in Autonomous Driving Research

Katalin Osz 1,2, Annie Rydström 2, Vaike Fors 1, Sarah Pink 3
Robert Broström

1 School of Information Technology, Halmstad University, 30118, Halmstad, Sweden
2 Volvo Car Group, 40531, Gothenburg, Sweden
3 Faculty of Information Technology, Monash Art, Design and Architecture, Monash
University, Melbourne, VIC 3800, Australia
{katalin, osz}@hh.se

Abstract. This article outlines a novel way of performing experimental “Wizard of Oz” (WOz) User Experience (UX) research that specifically targets driving in different levels of self-driving modes. The reasons for exploring the possibilities of combining experimental and ethnographic WOz-testing have been twofold. On the one hand, this mixed-method approach responds to a growing body of critique concerning how the WOz test is biased by the claim that it explores real-life behaviour in an experimental setting. On the other hand, our approach also meets the demands for innovative research methodologies that can contribute to deeper understandings of how to better evaluate and account for human expectations and experiences when automated technologies become integrated in everyday life contexts. This knowledge is inevitable for a broader understanding of the overall user experience and expectations of autonomous driving and, more specifically, building an interdisciplinary collaborative testing approach.

Keywords: autonomous cars, user experience, design anthropology, future technology, mixed-method approach

1 Introduction

In this article, we outline and demonstrate a new design research approach to creating insights about human expectations and experiences of a technology that does not yet fully exist - Autonomous Driving (AD). Significant advances in technology have made AD of vehicles a technological reality. Developing autonomous cars is technically challenging and to date the primary research focus regarding human behaviour in relation to AD cars has been on safety critical aspects, such as people’s ability to take over control from the automated car [1], [2], [3], [4], [5], mode awareness [6], overtrust [7, 8], or system transparency [9], with the intention of finding out how to provide the driver with optimal information through the user interfaces [10]. In the area of public and user acceptance of autonomous vehicles, a number of existing studies have focused on expectations through research into how to establish trust and mitigate resistance toward autonomous driving [11], [12], [13],
In this article, we present a mixed-method approach that brings together approaches from the field of User Experience (UX) and Human-Computer Interaction (HCI) research - experimental Wizard of Oz (WOz) testing - and ethnographic research informed by Design Anthropology. WOz is a research approach in which participants interact with a system that they believe to be real, but which is actually being operated or partially operated by an unseen person [16]. Design anthropology on the other hand offers a kind of human-centred approach, which can not only create ethnographic understandings of the intersection of technology and human life but is committed to produce an insightful, future-oriented way to account for human experience [17]. Our experience from automotive industry and existing literatures demonstrate that both approaches are now playing key roles in automotive design: WOz testing in experiment cars is an established and significant area of practice in automotive research which is essential for UX research, interaction design evaluation and safety assessments [5], [8], [18] and design anthropological and ethnographic research has recently taken on a high profile in AD research [19], [20], [21]. In this article, we report on the new knowledge and insights that the collaboration of these approaches produce and conclude how these approaches could be integrated in future empirical research to build more holistic user experience evaluations.

2 WOz testing in AD research

For an iterative UX design process there is a need to be able to put people in the right context when conducting evaluations. In the area of AD, evaluations need to be made in a realistic environment before fully functional AD cars are available. For this car simulators can be used, but to get a more realistic and ecologically valid test setup the WOz methodology is more promising. The WOz technique is an approach that has been used for evaluating user interfaces in various domains, from robotics [22] to mobile applications [23] and automotive industry [8, 18]. It is based on the idea of simulating a fully working technical system by a human operator – a wizard [24], and is used to gather data from users who believe they are interacting with an automated system. The WOz technique has been used in the automotive research community primarily for the design and study of automotive user interfaces, such as interfaces for driver assistance, information and entertainment [25], [26], [27], [28], [29]. Recently, it has been applied for evaluation of interactions with systems of higher level of automation [30] to gather data from users who believe they are experiencing and interacting with a highly-automated car. Compared with a real automated system, the WOz setup generally enables less constrained experiments - through use of improvisation. Also, it may enable more systematically constrained experiments - by cutting out the limitations of an automated system [31]. It does this in a way that is not reliant upon the development of new software and algorithms to control the vehicle, as is the case in real computer-operated systems [32]. Given its versatility, WOz is a good platform to examine interactions between humans and automated cars. However, the field of WOz testing, which is an important element of AD design research in the automotive industry [33], has tended to primarily remain attached to the specific psychological disciplinary orientation and set of analytic and research practices. Interactions, have mostly been focusing on momentary usability in a simulated setting to gather information about the nature of driver-car interaction but has lacked consistent, theoretical understanding of the concept of human experience [34].
3 Adding Design Ethnography to WOz testing

Design anthropology based ethnography combines methods from the social sciences and design to undertake in depth analysis of the real everyday contexts of product and service use. There are a number of renderings of the approach in existing practice and literatures. However, as developed here, its theoretical and methodological base is in design anthropological approaches [17], [35]. Design ethnography methods also often involve the use of video and intervention techniques that seek to use everyday environments and activities as probes to reveal new knowledge about everyday experience and imagination. These have been successfully developed in the field of design anthropology [19] in applied research in other technology fields in UK and Australia [36], [37], [38]. In AD research, experimental methodologies are still dominated by forms of testing that rely predominantly on engineering data [39]. However, the complementary perspectives that social science and particularly ethnographic insights offer in the area of user experience design have been increasingly acknowledged [40], [41], [42]. In our earlier research, we have used these methods to understand how participants experience and imagine AD features and automated services when driving [20] through in-car ethnographies in which we travelled with and video recorded participants in their cars and homes. The growing number of studies in this area demonstrate the vast, untapped potential for innovatively combining methodologies and bringing ethnography and the field of user experience together. In this article, we respond to this challenge by showing how we brought design ethnography into the practice of WOz experimental testing.

4 Filling the knowledge gaps: from ethnography of WOz testing to before-and-after WOz ethnography

The mixed methodological approach has been developed within the Human Expectations and Experiences of Autonomous Driving (HEAD) project (2016-2018). HEAD is a collaborative research project between Volvo Cars and Halmstad University in Sweden. The project is interdisciplinary with the aim to connect field testing, design ethnography and WOz user experience experimental testing to identify user expectations and experiences of self-driving cars. Collaboration between members of the Volvo Cars UX team and researchers from Halmstad University started in the summer of 2016, when researchers from Halmstad University started to visit, observe and participate in field experimental tests. Figure 1 below presents the chronological order of the collaborative tests. In the context of the research discussed here, we undertook ethnographic research with both participants in WOz tests and the WOz testers themselves. As a result, we established two collaborative links between
ethnography and WOz methodologies: first, ethnography of the WOz testing environment and second, ethnography as a before-and-after extension of WOz testing.

4.1 Ethnography of WOz testing

A series of interviews were conducted with the WOz testing team to generate a systematic overview of the context in which AD cars are prototyped, tested and test participants experience the WOz car. This involved creating a deep understanding of how the test environment has been set up, how the WOz car has been built (Figure 1) and how the team of testers’ knowledge, skills and learning have contributed to these processes over time. Interestingly, a systematic review of the test environment revealed some significant knowledge gaps. All interviewed testers, usability experts and interaction designers mentioned that novelty in user experience is brought about by unexpected, qualitative insights, however, the rigorously pre-planned test set-up does not allow much room for exploratory discoveries. There is conflict here because on the one hand, there is a growing need for novelty and user insights, but the way the WOz test is practiced, limits the potential of the methodology to be more exploratory. This methodological gap took us to the next step, where we needed to find new ways for the WOz testing to open up and sufficiently accommodate studying human expectations and experiences of AD.

![Fig. 2. The set-up of the WOz car, which is built on a Volvo XC90. In the set-up, the test participant is positioned in the front seat, and the driving and HMI (Human-Machine Interaction) wizards are positioned in the back seat using driving associated gears and built-in displays to simulate autonomy. A test leader can be positioned in the front or in the back, depending of its role in the study.](image)

4.2 Before-and-after-ethnography

In February 2018, Volvo Cars’ first WOz test on public roads took place. The aim was to test the usefulness of two prototype interfaces which were designed to indicate what current state the car was in, i.e. manual or fully autonomous. In the test eight study participants took part, all employed by Volvo Cars. All of them went through the same study procedure; they were asked to manually drive out of the Volvo Cars site to a motorway where they were guided by information provided on the interfaces (Figure 2). They were able to give over and take back control to and from the car. During autonomy mode, participants were able to engage in a task they had selected themselves, such as reading a paper, do work, writing e-mails on their smartphone or just relax and observe. Researchers from Halmstad University joined some of the WOz test drives. Before and after interviews with test participants took place directly before and after WOz tests. As summarised in Table 1, interviews and observations before the tests focused on commuting and driving routines, expectations and
imagination, mobilising images of how participants would imagine the AD experience to be like. Interviews after the WOz test reflected on the test environment and the overall AD experience including the tested interfaces.

Table 1. Focus areas of ethnographic research before and after WOz tests

<table>
<thead>
<tr>
<th>Before WOz</th>
<th>After WOz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commuting and driving routines</td>
<td>How tested interfaces would fit into existing driving routines</td>
</tr>
<tr>
<td>Automated features already in use</td>
<td>AD experience</td>
</tr>
<tr>
<td>Speculative scenarios</td>
<td>Trust</td>
</tr>
<tr>
<td>Learning about a new car</td>
<td>Reflections on what the AD car should learn from the driver</td>
</tr>
<tr>
<td>Time spent in the car</td>
<td>Communication between driver and the car</td>
</tr>
</tbody>
</table>

In the following section, we present an example, showing what the usability test looked like and what findings design ethnography added to the test. We then conclude the article by arguing that combining these approaches into a joint methodology have untapped potential to bring innovative insights and novelty into user experience research by capturing a more holistic understanding of AD experience.

Fig. 3. The route of the test. Participants were instructed to drive manually or using assisting functionality for longitudinal and lateral support until notified by the car that autonomous mode was available. When in autonomous mode the car asked the participant to take over the driving three times during the drive to explore how and when the participant discovered the information given by the prototype interfaces. The test took about 1 hour and 15 minutes for each participant.

5 From detached to holistic user insights: the AD motorway experience

In the WOz test an A/B-testing approach was used. A/B-testing is an integral part of an iterative design process, where two, or more, designs are tested on participants to determine which variation performs better. As described in the previous section, in this particular test two prototype interfaces for car status (manual mode or autonomous mode) were tested. The interfaces were based on different positions and
design of light indicators. The test also included usability testing, where the aim was to find usability problems in the designs. As commonly made in experimental testing, a set of pre-defined hypotheses were tested, such as “The light itself is sufficient to communicate AD status”. Questionnaires were also pre-designed, which restricted participants from freely reflecting on the test environment and AD experience. In the before and after ethnographic interviews, participants were able to discuss the overall AD test experience in relation to feelings (e.g. trust, mistrust or anxiety), their own commuting routines and idiosyncratic ways of driving. We found out more about how the tested usability concepts became part of the overall AD drive experience and why these design prototypes would or would not fit into participants’ everyday driving routines and idiosyncratic ways of driving. After the test drive, when we asked participants about the overall AD experience, the light-based control handover became a small part of the driving experience. Participants reflected more on trust, transparency and communication around the control handover, and were generally more concerned with how the WOz car indicated, kept distance, positioned in the lane, or reacted to other vehicles. All but one test participant expressed some sort of discomfort in relation to the way the WOz car was indicating, speeding over other cars and let other vehicles merge into the lane. Ultimately, most participants expressed concerns over the lack of information and communication based on which the WOz car made decisions.

Maria (46), a test participant and employee of Volvo Cars said: “There was a bus that wanted to go into my lane and the bus lane was coming to a stop and my car didn’t understand that. If I was driving myself I would have gone faster so the bus can come behind me or I would have slowed down so it could be in front of me. It is of course extremely difficult for the car to know that the bus wants to come into my lane. Even though the car was driving I felt that I need to check the road and be ready to intervene”. Mårten (42), another participant and employee expressed similar concerns: “At one moment, there was a car coming out at a faster speed. I would have chosen to wait for that car to pass because there weren’t any other cars behind that. And then, I would have changed because we had plenty of time. But this car chose to indicate and just change the lane in front of this car and that car had to change the lane. That didn’t make me trust in it. When I see something like that I think it wasn’t really a good behaviour”.

The design ethnography approach opened up the possibility for participants to freely reflect on their overall AD experience including a richer understanding of how the AD experience connects to aspirations, feelings and everyday commuting routines. Our findings started contouring novelty in the form of a rich variety and depth of expectations about what people would like the AD experience to be like. Participants did not only reflect on but envisioned what they would want the car in AD mode to learn from them as well as what the car should communicate under various road conditions and traffic situations. Therefore, combining ethnography with experiential WOz testing has helped us to draw out the idiosyncrasies of participant’s commuting and driving habits, making visible a range of personal foresights. Consequently, combining WOz testing with a design ethnographic approach offered a way to ground the development of AD experience in already existing driving and commuting habits. In addition, we observed two practical challenges that needed to be faced when building the combined test approach. First, it was more time consuming to plan and implement a shell test because the ethnographic research needed to be timed directly before and after the WOz tests. Not all research participants were able to free time during work hours, which is why the research team concluded that in-car ethnographies, where researchers drive with participants from their homes to work and back would be more suitable for future research. Second, long-term collaboration between UX and HCI researchers as well as anthropologists has been key to not only
build the test practice but effectively communicate and share research findings across various departments and disciplines.

6 Conclusion

In this article, we have reflected on the growing need to accommodate a larger set of analytical tools than what we have had until now to establish new, interdisciplin ary methodologies for studying expectations and experiences of AD. While hypothesis-based WOz testing seeks to gain insights from naturalistic behaviour, design ethnographic research is an exploratory approach that explicitly focuses on the encounter between the ethnographer and the research participants as the site of knowledge creation. We designed this study to investigate how these tensions can play out in a way that they bring value to the test practice. We concluded that adding a before-and-after ethnographic approach to WOz test drives opens up space to bring in the unexpected insights that often bring novelty by engaging and taking advantage of the emergent and improvised user insights of the test environment. These insights combine a diverse and unique set of needs, wants and expectations that AD UX design will need to fulfil, include and reflect on. In this article, we have also proposed that by adding ethnography to the WOZ testing practice, both the practice of building the WOz car and the evaluation of usability testing would be more grounded in human experience and expectations of everyday driving and commuting. Whereas WOz testing focuses specifically on the interaction between the vehicle and the user, in contrast, design ethnography compliments the interaction by insights where the driver is situated within the wider context of the driver-car-environment. Collaborative and extended inquiry in this field will continue with new ways to practice WOz testing in 2018, where before-and-after ethnography will not only extend the space for unexpected user insights, but will also focus on how to integrate these methodologies to address long-term effects of using AD cars in everyday life.

References


