How to Design Valid Simulator Studies for Investigating User Experience in Automated Driving – Review and Hands-On Considerations

Philipp Hock[⋆‡], **Johannes Kraus**[‡], **Franziska Babel**[‡], **Marcel Walch**[⋆], **Enrico Rukzio**[⋆] **Martin Baumann**[‡]
[⋆]Institute of Media Informatics
[‡]Institute of Psychology and Education, Dept. Human Factors
Ulm University, Germany
[philipp.hock, johannes.kraus, marcel.walch, franziska.babel, enrico.rukzio, martin.baumann]@uni-ulm.de

ABSTRACT
Simulator studies have been conducted in the automotive domain since the 1960s. Recently, automated driving studies have become more popular as real-world automated cars start to emerge but at this time not all levels of automation can be realized. A simulation does not entail all details of real driving, creating a realistic simulation experience - both on a psychological and physical level - proposes recurring challenges. These are among others: sample acquisition, simulator sickness, simulator training, interface design, take-over requests and secondary tasks in automated driving simulator studies. In this paper, we review existing literature and summarize important lessons from simulations in the domain of driving automation to provide considerations for studies investigating driver behavior in the age of highly automated driving.

Author Keywords
Automated driving; driving simulator; secondary task; simulator sickness; interface design; user studies

CCS Concepts
•**Human-centered computing** → **User studies;**

INTRODUCTION
Conducting studies that involve automated vehicles inherently involves an apparatus. The current state of maturity of automated vehicles cannot cover all capabilities of automation. Hence, the use of a real automated vehicle is not always possible. Furthermore, setups that involve dangerous situations, have to be done in some kind of simulation. As apparatus several technical systems are possible, such as a Wizard-Of-Oz automated vehicle [109] which is also applicable for virtual reality [30]. A more low fidelity approach is the use of video material to simulate an automated vehicle [55]. In this paper, we focus on the use of driving simulators as apparatus.

Driving simulators have been used in both academic and industrial research since the early 1960s, investigating driver behavior and interactions with the vehicle and the road [86]. Validity and reliability of driving simulator studies have been investigated frequently [5] and simulated driver behavior has been found to predict real-world driving to a considerable degree [47, 110, 94].

The advent of automated driving opens up a whole new domain of simulator studies. Searching the ACM Digital Library regarding the keywords *autonomous driving*, the publication count from 2010 to 2017 per year rises steadily from 22 to 85. Adding *driving simulator* to the filter, the publication count per year increases continuously from 3 up to 27. This indicates that research in automated or autonomous driving using simulators has become increasingly popular over the last 10 years. Experiments on user behavior in automated driving may focus on very different phenomena, for example system acceptance, trust, takeover behavior, cooperative strategies, usage of automation, etc. In consequence, new challenges, questions and problems regarding simulator studies for automated driving emerge.

In this paper, we review these challenges based on existing literature and the experiences from conducting driving simulator experiments for autonomous driving and those of fellow researchers in the field of driver-vehicle interaction in the age of (highly) automated driving. We derive basic considerations to conduct studies in a driving simulator and studies that investigate user behavior in automated vehicles with an emphasis on simulator studies on highly automated driving. The outline of this paper is structured along the decisions a researcher has to make during planning and conducting a driving simulator experiment: *acquisition and sample*, *briefing*, *responsibilities and takeover*, *secondary tasks*, *simulator training*, *simulator sickness*, *User Interface*, and *validity*.

AutomotiveUI '18, September 23–25, 2018, Toronto, ON, Canada

ISBN 978-1-4503-5946-7/18/09…15.00

DOI: https://doi.org/10.1145/3239060.3239066

PARTICIPANT ACQUISITION AND STUDY SAMPLE

In most HCI or simulator studies due to rather lengthy experimental procedures (information, training, etc.) and high costs of simulator operation [27, 7] small samples are tested. This increases the danger of sampling errors or biases and can enhance the effect of single outliers on the results in unpredicted ways.

Experience with automated driving influences participants' behavior to a considerable degree. For example, focusing on evaluations of user interfaces regarding automated driving, a tech-savvy and experienced sample may give different feedback than novices [106].

A very common pitfall in HCI research is that recruitment takes place at the labs, which mostly is a student population [7]. This results in a population with a narrow range in age, education, prior experience, motivation for participating, etc. Recruiting a narrow and small sample can lead to completely different findings when investigating attitudes towards automated driving compared to a sample of e.g. technological enthusiastic young drivers.

It is therefore possible, that the recruited sample will have low or no experience with autonomous cars as real world experience with those vehicles is scarce. This needs to be considered when interpreting the results of automated driving studies. The behaviour the participants show in the simulator is most often the initial reaction to the system and unlike the behaviour we would observe with a well-known system on the basis of daily interaction. Goodall [32] argued that future generations growing up with automated vehicles, will be more likely to rely on automation than today's drivers. We therefore recommend to report and discuss any possible implications of the very narrow sample. If, for example age, education, or affinity for technology mediates or moderates the results, a critical discussion is mandatory. Furthermore, we suggest to recruit a mixed sample including non-student participants.

Besides the participants' experience with autonomous vehicles or ADAS, participants' specific motivation to participate in the study can influence the study outcome. For example by advertising with *autonomous driving*, the participants' motivation may be different than when advertising with *driving simulator*. Especially when trust or acceptance is part of the research question (e.g. measuring the duration of time driven in autonomous mode), the participant's motivation to participate in the study is crucial for finding valid results. A participant who wants to test autonomous driving, may not be willing to disable the automation at all.

Therefore we suggest to consider not only where the advertising takes place, but also how the study's task and purpose is framed and explained. Above this, to control for these effects, the participant's study motivation could be assessed at the end of the study.

It is known that there are considerable influences of manifold factors on *trust* and *acceptance* of automated driving and consequently on automation use [58, 42, 97]. These influences can be separated in the categories: human-, automation or environment-related [36]. For recruitment and sampling procedures in driving simulator studies, especially the influences of individual differences and psychological traits need consideration [77, 81]. An overview on socio-demographic and personality-related variables identified in the literature is provided in Table 1.

User Characteristics	Source
Socio-demographics and experience	
Demographics	[36]
Age	
Gender	[42]
Subject matter expertise	[42, 24, 36]
Competency	[36]
Faith	[85, 70, 71]
Preexisting knowledge: expectations	[42]
Prior experience	[8, 36, 42, 58, 71]
History of interactions with the system	[8, 36, 42, 58, 71]
Training	[8, 36, 42, 58]
Personality	
Prospensity to trust	[36, 42, 58]
Neuroticism (Big Five)	[42]
Extraversion (Big Five)	[42, 10]
Dominance	[73]
Intuitive vs sensing personality	[42]
Locus of control	[2, 78, 88, 90, 91]
Sensation seeking	[43, 115, 78]
Attitudes towards the automation	[36, 42]
Comfort with robots	[36]
Self-confidence	[22, 42, 58, 36, 61]
Situational	
Mood	[42]
Attentional capacity	[42, 36]
Engagement	[36]

Table 1. Overview on socio-demographic and personality-related variables. The references provide a detailed discussion of these variables.

As a first example, in the area of personality traits, *locus of control* (LoC) [78, 91, 90] showed to be related with individual differences in trust in automation and automation use. Locus of control describes a person's tendency to which entity a cause of events is ascribed to – either own actions (internal LoC) or external factors beyond own control (external LoC) [2, 88]. Therefore, giving up control which is necessary for autonomous driving is especially difficult for people with an internal LoC. Thus, individuals with a higher tendency for an external LoC should be more positive about automated driving [90]. Indeed, external LoC has shown to influence the behavioural intention to use an automated vehicle [12] and led to increased trust in automation [91]. Examples for LoC scales are [54, 89]. As a second exemplary personality trait, *sensation seeking* has been shown to influence the usage of adaptive cruise control (ACC). Sensation seeking is defined as the need and search for intense experience and the willingness to take risk in order to reach these [115]. Individuals with higher sensation seeking tend to drive riskier even when ACC is engaged [78]. An example for a sensation seeking scales is [43]. Additionally, a negative relation of self-confidence and trust in automation was found in several studies [57, 16]. In studies examining the role of the Big Five Personality facets, extraversion was found to predict trust in automation [65] and neuroticism was negatively related to the agreement with an automated advice system [101]. In a recent study, significant correlations of trust and agreeableness and consciousness were reported [11]. Besides these rather general personality features, the specific attitude towards automated driving has been found to be a good indicator of how participants interact with the systems [65, 29].

This underlines the impact of several personality traits on the behavior of participants in simulator studies. Thus, they need to be considered and controlled for. Furthermore, prescreening for such predicting personality traits during recruitment of specific samples. At this point, it is unclear for many personality facts if they are general predictors for trust and automation use or if they only influence decision making under specific conditions. Taken together, the current state of research on the role of personality traits for the context of automated driving is scattered and in some regards inconclusive. Thus, more research in this area seems necessary. Furthermore, for simulator studies it seems recommendable to decide based on the research questions and on the basis of sample size and study setup to assess some selected traits that may influence the outcome of a specific study. The personality traits should be assessed with validated scales, which have to be chosen based on the available time and the respective study context. While a general recommendation seems not worthwhile, a short scale for sensation seeking can be found in [43] and a very brief measure for the Big Five in [33]. In order to test the participant's attitude towards automation prior to introducing the system, for instance the a priori acceptance of ADAS questionnaire [78] may be feasible. Note that the selection of assessed personality scales is specific for every study and should be based on hypotheses on how personality could affect participant behavior in the given study context.

BRIEFING

As long as autonomous vehicles are not part of our everyday mobility, participants will have little or no experience in the domain. Thus, participants' expectation and imagination of autonomous cars' abilities are diverse as they will in most cases not be based on facts but on word-of-mouth or emotions. To reduce the impact of this error variance to our experimental findings and for standardization and comparability, information prior to experimental interaction with the system should be used to adjust participant's knowledge [53, 38]. This includes explanation of capabilities, functionality and limitations of the automation. A crucial part of these instructions is the framing of the system, especially to explain the system's reliability [107, 52, 53]. Providing technical details about the automation could result in a mixed understanding of abilities among participants. For instance, telling participants that the car's sensors include radar might lead to mixed reactions in foggy situations because some participants could assume that radars are superior in such situations, other participants lack of such information [106]. This suggests that the briefing should contain the system's abilities and system boundaries rather than technical features to reduce the scope for interpretations. People have a bias towards automation and generally expect automated systems to outperform humans [22] and to work nearly flawless (perfect automation scheme) [66, 61]. In contrast to this, introducing an automation as not perfectly reliable may disappoint participants and thus lead them to rely even less on the system [16]. Trust develops dynamically depending on dispositional characteristics of the driver, situational factors and initial experience with the system (e.g. [58]). If a driver trusts the system and relies on it immediately, trust will develop differently than it would if he would not trust it at all. The initial level of trust influences the initial reliance strategy considerably. Therefore, experiences gathered in interaction with the system will always influence the course of development of trust into the automation resulting in an interdependent relation of trust and reliance [42]. The driver should use the automation optimally relying on the capabilities and knowing where the restraints of the system lie. If the driver is aware of system capabilities and restraints trust is well calibrated (calibrated trust), the introduction of the system should prevent drivers from overtrusting the system (using it in situations it is not capable of) or placing distrust in the automation (not using it although it would be capable) [58]. Taken together, if trust is a variable in the study, the system should not be introduced as flawless [57]. We propose that the instructions of the automation should be explained in detail in the publication for replication of the study.

Besides informing participants about system failures, the question arises if system failures should be demonstrated during the initial phase of the study. Witnessing a system failure can be expected to affect participants' trust towards the automation, even if it was made clear that the failure was intentionally implemented in this specific situation [45, 114]. We recommend to consider the influence of this exposure on the behavior and the results of the study and to decide individually whether to include a demonstration to system failures or to omit it.

Another important aspect in the instructions is to inform participants about the possibility of simulator sickness (see below for a thorough discussion of the phenomenon). The stronger participants expect simulator sickness, the more likely it will occur [51]. The way information is presented about simulation sickness, can amplify the experienced sickness because of a *framing effect* [103]. Accordingly, the introduction should be phrased in a way to avoid an increased simulator sickness. For instance, it should be avoided to tell participants that the simulation may cause vomiting.

SIMULATOR SICKNESS

Apart from the positive effects driving simulators pose for experimental research, they remain virtual environments which can potentially cause a perceptional conflict called simulation sickness. This includes symptoms like nausea, vertigo, sweating, headache and in the worst case sometimes vomiting [84, 49]. 80-90% of participants report simulator sickness related symptoms in simulator studies, that can affect about 5-30% of participants in such a way that they cannot finish the experiment [99]. There are different theories why simulator sickness and motion sickness occur. The popular sensory conflict theory explains simulator and motion sickness due to the discrepancy between the actual sensory information and the expected sensory information [83, 75, 6].

According to [37], characteristics that can worsen simulator sickness symptoms are: long exposure time, curves, turns, and higher speeds:

Cobb and colleagues [13] state that under one hour of exposure, simulator sickness symptoms steadily increase. After the one-hour mark is reached, symptoms return to a nominal level fifteen minutes later. Others [28, 50] suggested, that there might be be a positive and almost linear relationship between simulator sickness and exposure time. The best strategy against simulator sickness is exposure and adaption [98] but adaption itself is a highly individual process [56]. Participants are able to adapt to simulator sickness over multiple exposures in the simulator [98]. Some people will adapt over time while sickness is declined [56], but symptoms can also increase through repeated exposures [35]. About 5% of participants may not be able to adapt [39]. Road design also seems to affect simulator sickness. The road should be designed that there is a minimum of curves and if there are curves, they should not have small radii.

Surrounding conditions (like temperature) should be considered to mitigate simulator sickness [27]. Simulator sickness is higher in fixed-base driving simulators compared to moving-base simulators [15]. In fixed-based simulators, the feeling of acceleration and deceleration is absent. Therefore, feedback mechanisms to control speed are lacking [95]. Also, pre-screening of participants provides a perspective to control the problem. The Georgia Tech Simulator Sickness Screening Protocol (GTSSSP) [28] is one approach to pre-screen participants in order to predictively filter participants regarding increased risk for simulator sickness.

Despite occurring simulator sickness, some aspects of driving performance may not be influenced [37], but this mainly applies for speed and lane keeping behavior. But on the other hand, especially nausea can affect driving performance negatively. Loss of motivation, avoidance of tasks that are found disturbing, and possible distraction of the normal attention allocation process are possible negative effects [68]. The experimenter should also be aware that participants cannot decide whether they have to abort the experiment or not, therefore, we propose that the experimenter asks for their well being or the need for a pause as opposed to asking for aborting the experiment. Once the participant feels so sick that they need a pause, the likelihood of becoming sick again after the pause is high and it should be considered to abort the experiment [37]. Automated driving could increase simulator sickness compared to manual driving, which can lead to preventing the activation of the automation or engage in non-driving related tasks (NDRTs) [18]. Presumably because of the inability to anticipate the future motion path. Consequentially, implementing motion path cues could help to reduce simulator sickness [18]. The effects of lacking motion path cues also apply when performing secondary tasks. Therefore, preferring non-visual secondary tasks or locating displays in peripheral view of the road could reduce simulator sickness [18]. Therefore, we suggest to anticipate higher dropout rates compared to a manual driving simulator study. Severe symptoms of simulator sickness can influence the results, At the end of the study, it is common practice to ask participants for simulator sickness symptoms (e.g. [49]).

RESPONSIBILITIES AND TAKEOVER

To categorize the driver's level of responsibility, the SAE levels of automation can be used as an orientation [14]. When investigating NDRTs, execution of driving (level 1) must be on the side of the *system*, which is the case in SAE level 2 and above because NDRTs are only available when the human driver is currently not in control. The monitoring responsibility (level 2) and fallback entity (level 3) are the key factors when a takeover request (TOR) is triggered. We recommend to define the levels of automation of the system for the study to handle takeover situations properly (see the following section). However, naming the level of automation and definitions might not necessarily be in the instructions. It might be more prudent to explain the capabilities and boundaries of the system along with the responsibilities of the driver. Participants need not comprehend technical definitions but should know how to handle the automated driving feature, including malfunctions and takeovers [58, 22].

When TORs are implemented, it is important for participants to know how to react. A takeover is a system initiated control transition from automated drive to manual drive: automation-initiated driver control (AIDC), commonly due to a system boundary [60]. The participant's instructions should explain the differences between a system boundary and a malfunction.

A system boundary is well defined and predictable and is reliably detected by the system [31]. A malfunction occurs when something is not predictable, for example when a sensor is failing. For instance, a system boundary is reached when the vehicle is leaving the motorway if the automation can only drive on a motorway. The participant's instructions should explain all system boundaries and the required action, which most likely is a takeover. It is also important to explain the consequences if the participant fails to regain control over the car during the takeover. The consequences depend on the car's level of automation. The participant's instructions should also explain what happens when a system failure occurs [58, 22].

Other research has shown that the presence of motion cues [93], or traffic density [80] influences takeover performance. This suggests to, if feasible, aim for more realistic simulations in terms of perceived motion and consider the influence of simulated traffic on takeover performance. If not feasible, the limitations and implications should be discussed.

SECONDARY TASKS

One of the advantages of automated driving is the possibility to engage in NDRTs during driving. Such tasks can be any possible activity in the car, such as reading a newspaper or playing games on the phone. Depending on the car's level of automation, the driver must be capable to take over control of the car at all times and drive manually. The time to recover from a NDRT and to build up sufficient situational awareness is a factor very crucial for driving safety [25]. Some tasks may have a longer recovery time than others [69].

When the driver is involved in the driving task to a certain extend, which is the case in all levels of automation except for full automation (level 5), the NDRT becomes a distraction because the driving task may still become important at some point.

Driver distraction can be categorized in auditory, visual, cognitive, physical, technology-based and non-technology based distractions [112]. Physical distraction can further be distinguished according to the degree of interruptibility of the task [74]. An easily interrupted task is something that requires few or no action to stop the task and return to the driving task, such as reading a book or playing chess (without clock). Whereas hardly interruptible tasks require the user to do additionally steps to pause the task or involve some kind of punishment (e.g. watching movies then hit pause). Therefore, one should consider the factor of interruptibility of the secondary task as well as to discuss its influence on the performance.

Especially in the context of a research study, the priorities of the driving task and the secondary task can differ from real life as no real danger is given and losing a Tetris highscore may be more important for some participants than taking control of a simulated car.

Having a secondary task requires some judgment regarding the relative importance of both tasks – a task that is not trivial at all [23]. There are examples of other research [67] where performing well in the driving task was not incentivised but performing well in the secondary task has been encouraged. The validity of such studies is questionable and as stated in [67], such results only serve as worst case examples [67]. We discuss the influence of incentives in the Section *Validity and Behavior*.

As mentioned before, a reason for automation use can be gaining free time for NDRTs. The motivation to use automated driving features may depend on the possibilities to do something else during the ride in order to prevent underload [44, 113]. In this regard, the question arises what kind of secondary tasks can be provided in a driving simulator study. There are standardized tasks that can cause different kinds of distractions, such as SuRT (ISO/TS 14198) or n-back (e.g. [110, 80]). However, those tasks are usually not very entertaining and hence inappropriate to counteract boredom during automated driving.

There are several other approaches, some use selected literature to read or messages to write [21], some use smartphone games [104, 64], and other use movie trailers [108]. Another possibility to provide secondary tasks is that participants can bring their own tasks or use their private smartphones.

Participants may have a different degree of intrinsic motivation to engage in an otherwise unrewarded secondary task (e.g. some participants dislike playing Tetris whereas others love to). Providing other secondary tasks can have different influences on the usage of automation or even attention to TOR and the willingness to gain back control [26].

Letting participants chose their tasks for themselves, could create the highest external validity but also brings a highly uncontrolled factor into the study and consequently diminishes internal validity. In our literature research, we have found examples for all of the aforementioned techniques. All of these approaches have their own characteristics, drawbacks, advantages and possible consequences for the results. Therefore, we assume that there is not one perfect secondary task for simulator studies. In any case, motivational influences (possible differences of engagement) of the secondary task should be discussed when behavior is a matter of research.

The incentive to do NDRTs in a more naturalistic setting are intrinsic. Drivers may want to be productive or entertained. In a study, intrinsic motivations may not be present. To control the motivation of the NDRT, extrinsic motivation seems therefore adequate. The use of a controlled tasks, like watching movies combined with external incentives (e.g. provide higher payoffs for correctly answered questions regarding the movie (e.g. [108]) is one approach.

We suggest that the driving task should be introduced as the highest priority. To support this, accidents or even violations of traffic rules should influence the final payout negatively. This includes accidents and violations caused by the automation, because incentives serve as substitution for consequences, which are independent of the nature of the driver (human or machine). Besides the driving performance, the secondary task performance should be coupled with incentives to adapt the motivation in participation. Another approach is to provide tasks that participants have to do during the study but can decide when they want to complete the task. A participant can be given the choice do a secondary task during or after the ride to simulate productivity.

SIMULATOR TRAINING

Training sessions are common in HCI studies to avoid learning effects and to familiarize with a new system. Such training is also commonly a part of driving simulator experiments. This is especially important as the driving simulator responds differently to braking, accelerating and steering as a real vehicle. Main reasons for training include: (1) adoption, the reasons for a simulator training are (2) familiarization with the interior, like indicators and switches, (3) learning to handle study specific things like UI, and (4) reducing simulator sickness [19, 87]. (5) Adapting to transitioning between manual and automated driving, for example TOR (if such transitions occur in the study). (6) reduction of novelty, curiosity, and excitement and introduction to the system's behavior in specific [58, 42, 22].

In previous research, there is no uniform procedure for simulator training (e.g. time, distance, driving tasks) [95]. Some studies used a pre-defined period of time, a pre-defined distance, a self-assessment or a mixture [87]. The duration of the training is very diverse in previous research, reaching from two minutes [3] up to two days [76]. In between, some sessions lasted 5 minutes [102], 10 minutes [20], and 30 minutes [105].

When adaption is insufficient or even lacking at all, participants have to familiarize with the simulator during the actual experimental trials which influences the performance in the driving task and could bias the results of the experiment [95]. Spending more time in the simulator can improve adaption, while the probability for simulator sickness may rise [56, 50]. McGehee and colleagues [63] claim that a training benefit has reached its maximum after five minutes except for older drivers (65+) which need about eight minutes to adapt. The familiarization should cover the awareness and consequences of transitions and it should take the fear of executing such transitions [58]. If there are multiple ways to execute a transition (e.g. disabling the automation by a button, by braking, etc.) every method should be covered in the training session to prevent participants acting out of curiosity. A participant should not discover the features or behavior of the automation during the actual trials. For example, if participants do not know how the system performs when overtaking other cars, the participant should test this in the training and not in the test trial. A more complex and demanding track, requires a longer adaption period than a straight road [87].

Driving an automated car is a novel experience for most participants. Especially when investigating the usage of automated driving, the participants curiosity regarding automated driving might be a confounding factor. It is hard to eliminate this because participants may even be curious after having driven with the system for a while. Not only curiosity regarding automated driving, but also curiosity towards the simulator may exist. Therefore, not only the curiosity for automated driving, but also for manual driving should be considered.

After all six aspects of a simulator training have been considered, it is also important to check if the training had the required effect to a sufficient extend. For this, we propose three different approaches: (a) self-report of participants if they feel to have been provided with enough training [96, 87]. (b) evaluation of the participants' performance, e.g. lane deviation and stopping the training as soon as their performance reaches a certain threshold [95]. (c) evaluation of the correctness of the mental model of the participants by asking about specific situations or functions.

For each six aspects, we suggest different assessment strategies: For (1) *familiarization with the interior*: Asking participants if they have an understanding of the interior, as well as testing participant's knowledge for specific functions (e.g. indicator or light) might work well and we assume that the car interior and UI is understood very quick and intuitively.

For (2) *adapting to the vehicle dynamics*: performance indicators are measuring lane deviations, adherence to speed limits, and estimate breaking distances[27]. This can also be combined with asking participants if they feel comfortable to drive the car. The feeling of adaption can vary among participants from 10 [62] up to 30 minutes [79] and is an aspect that should be considered when scheduling participants.

For (3) *adapting to study specific features*: this is different for each study, therefore, we suggest to decide individually and explain the decisions appropriately.

For (4) *adapting to transitions*: if participants struggle with takeover situations, we recommend to include multiple transition changes within the test tracks.

For (5) *novelty*: it is hard to identify when participants are familiar with new situations. Therefore we suggest to ask the participant about their personal feeling about the system and if they want to further explore the (automated) system. The test drive should be continued until their impressions about the system remains constant.

For (6) *situations*: we suggest to run through every situation at least one time in order to demonstrate the system's behavior in each situation if possible. Participants should discover the abilities and the performance of the automation during the training. If this is not possible, the study design should include retesting of important situations.

UI

The user interface (UI) plays an important role in automated vehicles and influences the way automation is perceived and used. As the UI opens up the black box of automation and provides a large extent of the information used to built up the mental model of the automation, even slight modifications (e.g. in a HUD) could entail major consequences for the psychological reality of the driver interacting with an automated vehicle. This may lead to significant changes in variables like reaction times, subjective effort or decision making [34]. The appearance of the system influences trust towards the automation (e.g. [58, 42]). An interface that looks rather old or immature, can cause participants to depreciate the system's capabilities [42, 111, 58, 48].

To provide some guidance for designing user interfaces for automated driving, NHTSA has released the Federal Automated Vehicles Policy [1] in 2016 that includes minimum requirements an automated driving related human machine interface should inform the operator (driver) if the system is: (1) functioning properly, (2) currently engaged in automated driving mode, (3) currently unavailable for automated driving, (4) experiencing a malfunction, and (5) requesting control transition from automated driving to manual driving. Most driving simulators will not have any specific hardware indicating the status of the automation. Therefore, graphical interfaces play an important role in communicating the automation state.

We recommend being aware of the consequence of ambiguous UIs. The interface should communicate the automation state unambiguously [92]. Buttons for switching automation state like `Automation ON` might create confusion because the button might display the action or the current state. One solution to overcome this challenge could be dividing the UI into status and action components: the status display shows the current state of the automation (e.g. automation is: on | off | unavailable), the action button shows the future state of the automation (activate | deactivate). See Figure 1. Experi-

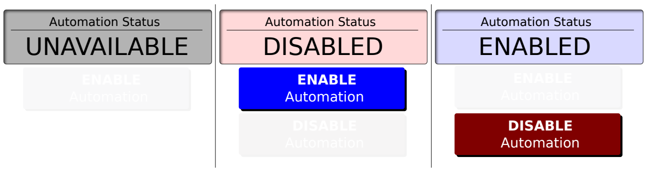

Figure 1. Example for an automated driving HMI. The Figure shows 3 different automation states from left to right: unavailable, disabled, and enabled. The automation status is color coded, unavailable is gray, disabled is red and enabled is blue. The button text encodes the action, not the current state to avoid ambiguities regarding automation state. The button for enabled and disabled are on different positions to avoid accidentally double pressing and switching states.

menters should also consider to implement multiple ways to disable and enable the automation, for example by pressing the acceleration or brake pedal. We recommend to evaluate the interface before the study and discuss possible confounding effects of the UI.

VALIDITY AND BEHAVIOR

When conducting simulator experiments, the inherently safe environment is often an advantage. Testing participant's behavior in hazardous situations or other situations in which participants would be in danger in the real world are almost only possible in simulations. Also in terms of experimental control a simulation provides crucial advantages over real world driving [27]. First, confounding variables such as weather, traffic and lighting can be standardized and controlled. Second, the driving situations can be easily repeatedly produced and study costs are considerably lower than real vehicle studies.

However, simulators cannot fully replicate real-life in every aspect [17]. Participants do not behave the same way, as they would in their own vehicle [27]. Although user reactions might be completely different in a simulator than in reality because nobody is exposed to real danger, a simulated crash may have a psychological effect on participants [27]. However, when hazardous situations are the scope of the investigation, it is often criticized that simulators can measure performance, but not behavior. This means it can be observed what participants *can do* but not what they *choose to do* [72, 82]. We currently face the problem that real automated vehicles cannot be used for any kind of automated driving studies due to restricted availability or indefensible dangers for the driving environment or participants. To date, the validity of automated driving simulator studies has not been fully investigated.

To address this problem, we first discuss several subcategories of validity: *Absolute validity*, describes the gap between results in a lab experiment and potential results in the real world (can be measured via the deviation of speed and lane position, task time, and visual attention) [110], whereas *relative validity* describes the direction of change. *Physical validity* describes the accuracy of a simulator's underlying visual and physical realism, for example rendering and calculating inertia. *Behavioral validity* describes the discrepancy between the actions and decisions during the experiments versus on-road behaviour [4, 110]. Both, behavioral and physical validity, are not necessarily aligned but it seems reasonable that a realistic simulation produces higher behavioral validity [9]. The term *performance validity* describes parameters like speed and lane keeping.

We assume that absolute validity in automated simulator studies is hard to achieve and results claiming absolute validity may be questionable when decision making or behavioral aspects are measured. However, we further assume that behavioral validity can be investigated in a simulated automated vehicle [47, 110, 94].

It is known practice to replace components that have been eliminated by the simulation through instructions and incentives. Improving relative behavioral validity can be achieved by incentives, like a reward/penalty system [27]. Although the consequences for the participant cannot be the same as in reality, the participant's reaction in the simulator to a certain event should be as close as possible to the same event in real life. Achieving this is not trivial. In reality, the consequences for a failing automation reach from nothing, over cost up to death. Besides the immediate consequences, there are also consequences regarding the attitude towards the automation itself that is likely to be influenced. When the automation fails, drivers may be concerned about the automation's capabilities for a long time. Our observations in simulations however show that trust in the automation rises very quick after a failure happens or a takeover request occurs, which is in line with the literature [40, 77, 91].

Incentives for driving simulators can be distinguished in the following categories [27]: (1) Consequential, which can lead to real world consequences. For example when simulations are used to measure the subject's ability to perform a tasks, like pilots failing in the simulation leading to consequences in real life. Here the simulation has consequences regarding the permission to operate on machines in reality. This does not apply for simulator studies. (2) Intrinsic incentives, that comes from the participant's willingness to perform well in a specific task, for example because of a given challenge. (3) Extrinsic incentives, like monetary rewards. It is considered ethical to use deception at the beginning of a study and make participants believe that it will impact their compensation/reward, but then give all participants the same reward afterwards anyway.

In driving simulations practitioners rely mainly on intrinsic and extrinsic incentives. One approach to add consequential incentives could be to add raffle for an additional monetary reward, like a voucher that one of the participants can win. Then lower the chances to win the voucher if participants are involved in accidents or similar.

There are several approaches for incentives in simulator studies. Stein and colleagues [100] describe a monetary reward/penalty system where faster completion times were rewarded, speeding was associated with speeding tickets resulting in monetary penalties. The ticket was not always fined but followed a parameterized probabilistic algorithm. Crashes influenced both, completion time and money. The effect can be enhanced by adding noises on specific events, like sirens for tickets and breaking glass as well as screeching tires for a crash [59, 27]. The idea of completion time can also be adopted to automated driving. For instance, the automation can be instructed as driving more ecologically which pays off as rewards in the end of the experiment to encourage using automated driving. On the other hand, the automation may drive very defensive and slow compared to an average human driver to encourage manual driving [41]. In this study, the automation was slowed down by a vehicle ahead. When participants canceled the automation and took over the car in front, they could drive faster. This behavior was rewarded because their payment was linked linearly to the length driven on the track. When participants drove manually and as fast as allowed and possible, they would get 9€, when the automation drove the entire time, they would get 6€. We assume that this effect could be enhanced when instructing that participants will get the full amount (e.g. 9€) and that a certain behavior (e.g. driving a fewer distance, causing crashed) result in subtractions of the end value because according to the prospect theory [46], 'losses loom larger than gains'. The incentive does not necessarily have to consist of money.

CHECKLIST

Stepping through the process of automated driving simulator studies, we identified eight stages where we found typical challenges to overcome when conducting studies. We summarized all considerations from the text above into the checklist below.

SAMPLE

- Recruit a mixed sample, report and discuss the sample and limitations.
- Consider and assess psychological traits and socio-demographics (e.g. LoC, sensation seeking and attitudes towards the automation) (possibly pre-screen).
- Consider selection bias. Avoid advertising that addresses the curiosity or motivation of participants.

BRIEFING

- Exclude technical information from the briefing.
- Describe system's capabilities and system boundaries instead.
- Decide whether to include a demonstration of system failures.
- Refrain from introducing the system as flawless, if trust is of interest in the study.
- Explain the instructions of the automation in any publication to facilitate replication.

SIMULATOR SICKNESS

- Avoid long exposure times, narrow curves, sharp turns, and high speeds.
- Pre-screen participants (e.g. GTSSSP) to exclude participants experiencing severe symptoms of simulator sickness.
- At the end of the study, assess simulator sickness (e.g. SSQ) to avoid confounded results.
- Ask for participants' well-being and need for breaks.
- Stop the experiment if the participant feels severe symptoms of simulator sickness.
- Implement motion path cues.
- Consider framing effects when briefing simulator sickness.

TAKEOVER

- Explain the differences between a system boundary and a malfunction and the desired action if a TOR occurs.
- Aim for realistic motion simulation.
- Consider simulated traffic as an influence on TOR performance.

SECONDARY TASK

- We recommend to discuss motivational influences (possible differences of engagement) of the secondary task when behavior is a matter of research.
- Control the motivation for the secondary task with external incentives.
- Introduce the driving task as the primary task.
- Preferring non-visual secondary tasks or locate displays in peripheral view of the road to reduce simulator sickness.

SIMULATOR TRAINING

- Simulator training should cover: adoption, familiarization with the interior and the study setup, transitioning between manual and automated driving, reduce simulator sickness, and novelty effects.
- Check training effects: self-reports from participants, performance evaluation, and participant's knowledge test.

USER INTERFACE

- Avoid ambiguous UIs for example by separating the action buttons from the automation states.
- Implement multiple ways to enable and disable the automation.

VALIDITY

- Introduce a penalty/reward system as consequences for driving behaviour.
- Let accidents/violation of traffic rules influence the final payout negatively.

CONCLUSION

In this paper, we provided insights and practical suggestion for conducting simulator user studies in automated driving. We investigated common aspects of recent automated driving studies.

We see our listings and suggestion as useful hints and not as a complete standard. We discussed and examined different subtopics of automated driving simulator studies. We mainly focus on a literature research and also on self reports and observations. In order to extend our considerations into guidelines or standards, our observations need to be confirmed with user studies. One question that needs further investigation will be the general validity of automated driving simulator experiments because of the safe environment. Another topic that needs further investigations is the use of secondary tasks and NDRTs. There are different approaches what participants can do while driving automated. Our observations provide useful insights and a concrete checklist of aspects to consider when planning and conducting simulator experiments in the context of automated driving.

ACKNOWLEDGMENTS

This work was funded by the Carl-Zeiss-Scholarship for PhD students.

REFERENCES

1. National Highway Traffic Safety Administration and others. 2016. Federal automated vehicles policy. *Accelerating the next revolution in roadway safety. DOT HS* 812 (2016), 329.
2. Icek Ajzen. 2002. Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of applied social psychology* 32, 4 (2002), 665–683.
3. Peter H Baas, Samuel G Charlton, and Gary T Bastin. 2000. Survey of New Zealand truck driver fatigue and fitness for duty. *Transportation research part F: traffic psychology and behaviour* 3, 4 (2000), 185–193.
4. Gerard J Blaauw. 1982. Driving experience and task demands in simulator and instrumented car: a validation study. *Human Factors* 24, 4 (1982), 473–486.
5. Evi Blana. 1996. Driving Simulator Validation Studies: A Literature Review. (1996).
6. Willem Bles, Jelte E Bos, Bernd de Graaf, Eric Groen, and Alexander H Wertheim. 1998. Motion sickness: only one provocative conflict? *Brain research bulletin* 47, 5 (1998), 481–487.
7. Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 981–992. DOI: http://dx.doi.org/10.1145/2858036.2858498
8. Erika N Carlson and R Michael Furr. 2013. Resolution of meta-accuracy: Should people trust their beliefs about how others see them? *Social Psychological and Personality Science* 4, 4 (2013), 419–426.
9. Oliver Carsten and A Hamish Jamson. 2011. Driving simulators as research tools in traffic psychology. *Handbook of traffic psychology* 1 (2011), 87–96.
10. Justine Cassell and Timothy Bickmore. 2003. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modeling and User-Adapted Interaction* 13, 1 (2003), 89–132.
11. Shih-Yi Chien, Michael Lewis, Katia Sycara, Jyi-Shane Liu, and Asiye Kumru. 2016. Influence of cultural factors in dynamic trust in automation. In *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*. IEEE, 002884–002889.
12. Jong Kyu Choi and Yong Gu Ji. 2015. Investigating the Importance of Trust on Adopting an Autonomous Vehicle. *International Journal of Human-Computer Interaction* 31, 10 (2015), 692–702. DOI: http://dx.doi.org/10.1080/10447318.2015.1070549
13. Sue VG Cobb, Sarah Nichols, Amanda Ramsey, and John R Wilson. 1999. Virtual reality-induced symptoms and effects (VRISE). *Presence: teleoperators and virtual environments* 8, 2 (1999), 169–186.
14. SAE On-Road Automated Vehicle Standards Committee and others. 2016. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. *SAE Standard J3016* (2016), 01–30.
15. R Curry, B Artz, L Cathey, P Grant, and J Greenberg. 2002. Kennedy ssq results: fixed-vs motion-based FORD simulators. In *Proceedings of DSC*. 289–300.
16. Peter de Vries, Cees Midden, and Don Bouwhuis. 2003. The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human-Computer Studies* 58, 6 (2003), 719–735.
17. J De Winter, P Van Leuween, and P Happee. 2012. Advantages and disadvantages of driving simulators: A discussion. In *Proceedings of Measuring Behavior*. 47–50.
18. Cyriel Diels and Jelte E Bos. 2016. Self-driving carsickness. *Applied ergonomics* 53 (2016), 374–382.
19. Joshua E Domeyer, Nicholas D Cassavaugh, and Richard W Backs. 2013. The use of adaptation to reduce simulator sickness in driving assessment and research. *Accident Analysis & Prevention* 53 (2013), 127–132.
20. Lisa Dorn and David Barker. 2005. The effects of driver training on simulated driving performance. *Accident Analysis & Prevention* 37, 1 (2005), 63–69.
21. Frank A Drews, Hina Yazdani, Celeste N Godfrey, Joel M Cooper, and David L Strayer. 2009. Text messaging during simulated driving. *Human factors* 51, 5 (2009), 762–770.
22. Mary T Dzindolet, Scott A Peterson, Regina A Pomranky, Linda G Pierce, and Hall P Beck. 2003. The role of trust in automation reliance. *International Journal of Human-Computer Studies* 58, 6 (2003), 697–718.
23. Ward Edwards. 1961. Costs and payoffs are instructions. *Psychological Review* 68, 4 (1961), 275.
24. Fredrick Ekman, Mikael Johansson, and Jana L Sochor. 2016. Creating appropriate trust for autonomous vehicle systems: A framework for HMI Design. In *Proceedings of the 95th Annual Meeting of the Transportation Research Board, Washington, DC January 10-14, 2016*.
25. Mica R Endsley. 1995. Toward a theory of situation awareness in dynamic systems. *Human factors* 37, 1 (1995), 32–64.

26. Alexander Eriksson and Neville A Stanton. 2017. Takeover time in highly automated vehicles: noncritical transitions to and from manual control. *Human factors* 59, 4 (2017), 689–705.

27. Donald L Fisher, Matthew Rizzo, Jeffrey Caird, and John D Lee. 2011. *Handbook of driving simulation for engineering, medicine, and psychology*. CRC Press.

28. Thomas M Gable and Bruce N Walker. 2013. *Georgia Tech Simulator Sickness Screening Protocol*. Technical Report. Georgia Institute of Technology.

29. David Garcia, Christine Kreutzer, Karla Badillo-Urquiola, and Mustapha Mouloua. 2015. Measuring Trust of Autonomous Vehicles: A Development and Validation Study. In *International Conference on Human-Computer Interaction*. Springer, 610–615.

30. David Goedicke, Jamy Li, Vanessa Evers, and Wendy Ju. 2018. VR-OOM: Virtual Reality On-rOad Driving siMulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 165, 11 pages. DOI: http://dx.doi.org/10.1145/3173574.3173739

31. Christian Gold, Daniel DambÃűck, Lutz Lorenz, and Klaus Bengler. 2013. âĂIJTake over!âĂİ How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 57, 1 (2013), 1938–1942. DOI: http://dx.doi.org/10.1177/1541931213571433

32. Noah Goodall. 2014. Ethical decision making during automated vehicle crashes. *Transportation Research Record: Journal of the Transportation Research Board* 2424 (2014), 58–65.

33. Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528.

34. Martin Götze, Christian Schweiger, Johannes Eisner, and Klaus Bengler. 2015. Comparison of an old and a new head-up display design concept for urban driving. (2015).

35. Ashton Graybiel and James R Lackner. 1983. Motion sickness: acquisition and retention of adaptation effects compared in three motion environments. *Aviation, space, and environmental medicine* (1983).

36. Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors* 53, 5 (2011), 517–527.

37. Arne Helland, Stian Lydersen, Lone-Eirin Lervåg, Gunnar D Jenssen, Jørg Mørland, and Lars Slørdal. 2016. Driving simulator sickness: Impact on driving performance, influence of blood alcohol concentration, and effect of repeated simulator exposures. *Accident Analysis & Prevention* 94 (2016), 180–187.

38. Tove Helldin, Göran Falkman, Maria Riveiro, and Staffan Davidsson. 2013. Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. In *Proceedings of the 5th international conference on automotive user interfaces and interactive vehicular applications*. ACM, 210–217.

39. K Helling and M Westhofen. 1994. Experimental studies of motion sickness on board of a research ship. *Hno* 42, 4 (1994), 214–219.

40. Sebastian Hergeth, Lutz Lorenz, Josef F Krems, and Lars Toenert. 2015. Effects of take-over requests and cultural background on automation trust in highly automated driving. (2015).

41. Philipp Hock, Johannes Kraus, Marcel Walch, Nina Lang, and Martin Baumann. 2016. Elaborating Feedback Strategies for Maintaining Automation in Highly Automated Driving. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (Automotive'UI 16)*. ACM, New York, NY, USA, 105–112. DOI: http://dx.doi.org/10.1145/3003715.3005414

42. Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors* 57, 3 (2015), 407–434.

43. Rick H Hoyle, Michael T Stephenson, Philip Palmgreen, Elizabeth Pugzles Lorch, and R Lewis Donohew. 2002. Reliability and validity of a brief measure of sensation seeking. *Personality and individual differences* 32, 3 (2002), 401–414.

44. A Hamish Jamson, Natasha Merat, Oliver MJ Carsten, and Frank CH Lai. 2013. Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation research part C: emerging technologies* 30 (2013), 116–125.

45. Daniel Kahneman and Patrick Egan. 2011. *Thinking, fast and slow*. Vol. 1. Farrar, Straus and Giroux New York.

46. Daniel Kahneman and Amos Tversky. 2013. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, 99–127.

47. Nico Kaptein, Jan Theeuwes, and Richard Van Der Horst. 1996. Driving simulator validity: Some considerations. *Transportation Research Record: Journal of the Transportation Research Board* 1550 (1996), 30–36.

48. KAPJ Karvonen and J Parkkinen. 2001. Signs of trust: A semiotic study of trust formation in the web. In *First international conference on universal Access in human-computer interaction*, Vol. 1. 1076–1080.

49. Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220.

50. Robert S Kennedy, Kay M Stanney, and William P Dunlap. 2000. Duration and exposure to virtual environments: sickness curves during and across sessions. *Presence: Teleoperators and virtual environments* 9, 5 (2000), 463–472.

51. Behrang Keshavarz and Heiko Hecht. 2011. Validating an efficient method to quantify motion sickness. *Human factors* 53, 4 (2011), 415–426.

52. Jeamin Koo, Jungsuk Kwac, Wendy Ju, Martin Steinert, Larry Leifer, and Clifford Nass. 2015. Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 9, 4 (2015), 269–275.

53. Moritz Körber, Eva Baseler, and Klaus Bengler. 2018. Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied ergonomics* 66 (2018), 18–31.

54. Anastassiya Kovaleva, Constanze Beierlein, Christoph J. Kemper, and Beatrice Rammstedt. 2012. Eine Kurzskala zur Messung von Kontrollüberzeugung: Die Skala Internale-Externale-Kontrollüberzeugung-4 (IE-4). *GESIS-Working Papers* 19 (2012), 1–26. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis

55. Johannes Maria Kraus, Jessica Sturn, Julian Elias Reiser, and Martin Baumann. 2015. Anthropomorphic Agents, Transparent Automation and Driver Personality: Towards an Integrative Multi-level Model of Determinants for Effective Driver-vehicle Cooperation in Highly Automated Vehicles. In *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '15)*. ACM, New York, NY, USA, 8–13. DOI: http://dx.doi.org/10.1145/2809730.2809738

56. James R Lackner. 2014. Motion sickness: more than nausea and vomiting. *Experimental brain research* 232, 8 (2014), 2493–2510.

57. John D Lee and Neville Moray. 1994. Trust, self-confidence, and operators' adaptation to automation. *International journal of human-computer studies* 40, 1 (1994), 153–184.

58. John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

59. Mark R Lehto, JD Papastavrou, Thomas A Ranney, and Leigh Ann Simmons. 2000. An experimental comparison of conservative versus optimal collision avoidance warning system thresholds. *Safety Science* 36, 3 (2000), 185–209.

60. Zhenji Lu and Joost CF de Winter. 2015. A review and framework of control authority transitions in automated driving. *Procedia Manufacturing* 3 (2015), 2510–2517.

61. Poornima Madhavan and Douglas A Wiegmann. 2007. Similarities and differences between human–human and human–automation trust: an integrative review. *Theoretical Issues in Ergonomics Science* 8, 4 (2007), 277–301.

62. Deborah McAvoy, Kerrie Schattler, and Tapan Datta. 2007. Driving simulator validation for nighttime construction work zone devices. *Transportation Research Record: Journal of the Transportation Research Board* 2015 (2007), 55–63.

63. Daniel V McGehee, John D Lee, Matthew Rizzo, and Kirk Bateman. 2001. Examination of older driver steering adaptation on a high performance driving simulator. (2001).

64. Vivien Melcher, Stefan Rauh, Frederik Diederichs, Harald Widlroither, and Wilhelm Bauer. 2015. Take-over requests for automated driving. *Procedia Manufacturing* 3 (2015), 2867–2873.

65. Stephanie M Merritt and Daniel R Ilgen. 2008. Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors* 50, 2 (2008), 194–210.

66. Stephanie M Merritt, Jennifer L Unnerstall, Deborah Lee, and Kelli Huber. 2015. Measuring individual differences in the perfect automation schema. *Human factors* 57, 5 (2015), 740–753.

67. Brian Mok, Mishel Johns, David Miller, and Wendy Ju. 2017. Tunneled In: Drivers with Active Secondary Tasks Need More Time to Transition from Automation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 2840–2844. DOI: http://dx.doi.org/10.1145/3025453.3025713

68. MA Mollenhauer, RA Romano, and B Brumm. 2006. The evaluation of a motion base driving simulator in a CAVE at TACOM. In *Transformational Science And Technology For The Current And Future Force: (With CD-ROM)*. World Scientific, 323–330.

69. Christopher A Monk, Deborah A Boehm-Davis, George Mason, and J Gregory Trafton. 2004. Recovering from interruptions: Implications for driver distraction research. *Human factors* 46, 4 (2004), 650–663.

70. Bonnie M Muir. 1994. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics* 37, 11 (1994), 1905–1922.

71. Bonnie M Muir and Neville Moray. 1996. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics* 39, 3 (1996), 429–460.

72. Risto Näätänen and Heikki Summala. 1976. Road-user behaviour and traffic accidents. *Publication of: North-Holland Publishing Company* (1976).

73. Clifford Nass, Youngme Moon, BJ Fogg, Byron Reeves, and D Christopher Dryer. 1995. Can computer personalities be human personalities? *International Journal of Human-Computer Studies* 43, 2 (1995), 223–239.

74. Y Ian Noy, Tracy L Lemoine, Christopher Klachan, and Peter C Burns. 2004. Task interruptability and duration as measures of visual distraction. *Applied Ergonomics* 35, 3 (2004), 207–213.

75. Charles M Oman. 1990. Motion sickness: a synthesis and evaluation of the sensory conflict theory. *Canadian journal of physiology and pharmacology* 68, 2 (1990), 294–303.

76. Timothy O'Neill, Gerald Krueger, Susan Van Hemel, Adam McGowan, and William Rogers. 1999. Effects of cargo loading and unloading on truck driver alertness. *Transportation Research Record: Journal of the Transportation Research Board* 1686 (1999), 42–48.

77. Raja Parasuraman and Victor Riley. 1997. Humans and automation: Use, misuse, disuse, abuse. *Human factors* 39, 2 (1997), 230–253.

78. William Payre, Julien Cestac, and Patricia Delhomme. 2014. Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation research part F: traffic psychology and behaviour* 27 (2014), 252–263.

79. Eli Peli, A Bowers, A Mandel, K Higgins, R Goldstein, and L Bobrow. 2005. Design for simulator performance evaluations of driving with vision impairments and visual aids. *Transportation Research Record: Journal of the Transportation Research Board* 1937 (2005), 128–135.

80. Jonas Radlmayr, Christian Gold, Lutz Lorenz, Mehdi Farid, and Klaus Bengler. 2014. How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 58. Sage Publications Sage CA: Los Angeles, CA, 2063–2067.

81. Bako Rajaonah, Nicolas Tricot, Françoise Anceaux, and Patrick Millot. 2008. The role of intervening variables in driver–ACC cooperation. *International Journal of Human-Computer Studies* 66, 3 (2008), 185–197.

82. Thomas Ranney, Ginger S Watson, Elizabeth N Mazzae, Yiannis E Papelis, Omar Ahmad, and Judith R Wightman. 2005. *Examination of the Distraction Effects of Wireless Phone Interfaces Using the National Advanced Driving Simulator-Final Report on a Freeway Study*. Technical Report.

83. James T Reason. 1978. Motion sickness adaptation: a neural mismatch model. *Journal of the Royal Society of Medicine* 71, 11 (1978), 819.

84. René Reinhard, Hans M Rutrecht, Patricia Hengstenberg, Ender Tutulmaz, Britta Geissler, Heiko Hecht, and Axel Muttray. 2017. The best way to assess visually induced motion sickness in a fixed-base driving simulator. *Transportation research part F: traffic psychology and behaviour* 48 (2017), 74–88.

85. John K Rempel, John G Holmes, and Mark P Zanna. 1985. Trust in close relationships. *Journal of personality and social psychology* 49, 1 (1985), 95.

86. King M Roberts. 1980. FHWA HIGHWAY DRIVING SIMULATOR. *Public Roads* 44, HS-031 069 (1980).

87. Adi Ronen and Nitzan Yair. 2013. The adaptation period to a driving simulator. *Transportation research part F: traffic psychology and behaviour* 18 (2013), 94–106.

88. J. B. Rotter. 1966a. Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs* 80, 1 (1966), 1–28. DOI: http://dx.doi.org/10.1037/h0092976

89. Julian B Rotter. 1966b. Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied* 80, 1 (1966), 1.

90. Christina Rudin-Brown and Y Ian Noy. 2002. Investigation of behavioral adaptation to lane departure warnings. *Transportation Research Record: Journal of the Transportation Research Board* 1803 (2002), 30–37. DOI: http://dx.doi.org/https://doi.org/10.3141/1803-05

91. Christina M. Rudin-Brown and Heather A. Parker. 2004. Behavioural adaptation to adaptive cruise control (ACC): Implications for preventive strategies. *Transportation Research Part F: Traffic Psychology and Behaviour* 7, 2 (2004), 59–76. DOI: http://dx.doi.org/10.1016/j.trf.2004.02.001

92. John Rushby. 2002. Using model checking to help discover mode confusions and other automation surprises. *Reliability Engineering & System Safety* 75, 2 (2002), 167–177.

93. Shadan Sadeghian Borojeni, Susanne CJ Boll, Wilko Heuten, Heinrich H Bülthoff, and Lewis Chuang. 2018. Feel the Movement: Real Motion Influences Responses to Take-over Requests in Highly Automated Vehicles. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 246.

94. Saeed Sahami. 2011. *Modeling adaptation behavior to driving simulators and effect of experimental practice on research validity*. Ph.D. Dissertation. University of British Columbia.

95. Saeed Sahami and Tarek Sayed. 2013. How drivers adapt to drive in driving simulator, and what is the impact of practice scenario on the research? *Transportation research part F: traffic psychology and behaviour* 16 (2013), 41–52.

96. Nikhil Ravindra Sarwate. 2015. Skill Retention for Driving Simulation Experiments. (2015).

97. Kristin E Schaefer, Jessie YC Chen, James L Szalma, and Peter A Hancock. 2016. A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human factors* 58, 3 (2016), 377–400.

98. Frank Schmäl. 2013. Neuronal mechanisms and the treatment of motion sickness. *Pharmacology* 91, 3-4 (2013), 229–241.

99. Kay M Stanney, Ronald R Mourant, and Robert S Kennedy. 1998. Human factors issues in virtual environments: A review of the literature. *Presence: Teleoperators and Virtual Environments* 7, 4 (1998), 327–351.

100. Anthony C Stein, R Wade Allen, and Stephen H Schwartz. 1978. Use of reward-penalty structures in human experimentation. (1978).

101. James L Szalma and Grant S Taylor. 2011. Individual differences in response to automation: The five factor model of personality. *Journal of Experimental Psychology: Applied* 17, 2 (2011), 71.

102. Pierre Thiffault and Jacques Bergeron. 2003. Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis & Prevention* 35, 3 (2003), 381–391.

103. Amos Tversky and Daniel Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211, 4481 (1981), 453–458.

104. Unfallforschung der Versicherer – GDV. 2016. Takeover times in highly automated driving. https://udv.de/en/publications/compact-accident-research/takeover-times-highly-automated-driving. (2016). Accessed: 25-January-2018.

105. Wim van Winsum, Dick de Waard, and Karel A Brookhuis. 1999. Lane change manoeuvres and safety margins. *Transportation Research Part F: Traffic Psychology and Behaviour* 2, 3 (1999), 139–149.

106. Viswanath Venkatesh and Hillol Bala. 2008. Technology acceptance model 3 and a research agenda on interventions. *Decision sciences* 39, 2 (2008), 273–315.

107. Frank MF Verberne, Jaap Ham, and Cees JH Midden. 2012. Trust in smart systems: Sharing driving goals and giving information to increase trustworthiness and acceptability of smart systems in cars. *Human factors* 54, 5 (2012), 799–810.

108. Marcel Walch, Tobias Sieber, Philipp Hock, Martin Baumann, and Michael Weber. 2016. Towards Cooperative Driving: Involving the Driver in an Autonomous Vehicle's Decision Making. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (Automotive'UI 16)*. ACM, New York, NY, USA, 261–268. DOI: http://dx.doi.org/10.1145/3003715.3005458

109. Peter Wang, Srinath Sibi, Brian Mok, and Wendy Ju. 2017. Marionette: Enabling On-Road Wizard-of-Oz Autonomous Driving Studies. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. ACM, New York, NY, USA, 234–243. DOI: http://dx.doi.org/10.1145/2909824.3020256

110. Ying Wang, Bruce Mehler, Bryan Reimer, Vincent Lammers, Lisa A D'Ambrosio, and Joseph F Coughlin. 2010. The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing. *Ergonomics* 53, 3 (2010), 404–420.

111. Alona Weinstock, Tal Oron-Gilad, and Yisrael Parmet. 2012. The effect of system aesthetics on trust, cooperation, satisfaction and annoyance in an imperfect automated system. *Work* 41, Supplement 1 (2012), 258–265.

112. Kristie Young, Michael Regan, and M Hammer. 2007. Driver distraction: A review of the literature. *Distracted driving* (2007), 379–405.

113. Mark S Young and Neville A Stanton. 2002. Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Human factors* 44, 3 (2002), 365–375.

114. Robert B Zajonc. 1968. Attitudinal effects of mere exposure. *Journal of personality and social psychology* 9, 2p2 (1968), 1.

115. Marvin Zuckerman. 1994. *Behavioral expressions and biosocial bases of sensation seeking*. Cambridge university press.