

Article

# Where We Come from and Where We Are Going: A Systematic Review of Human Factors Research in Driving Automation

Anna-Katharina Frison <sup>1,2,3,†</sup>, Yannick Forster <sup>4,\*,†</sup>, Philipp Wintersberger <sup>1,2,†</sup>, Viktoria Geisel <sup>5</sup>  
and Andreas Riemer <sup>1,2</sup>

<sup>1</sup> Technische Hochschule Ingolstadt, Esplanade 10, 85049 Ingolstadt, Germany; anna-katharina.frison@frisux.com (A.-K.F.); philipp.wintersberger@carissma.eu (P.W.); andreas.riemer@thi.de (A.R.)

<sup>2</sup> Institute for Pervasive Computing, Johannes Kepler University, Altenbergerstr. 69, 4040 Linz, Austria

<sup>3</sup> frisUX, Stocketweg 5, 83355 Grabenstaett, Germany

<sup>4</sup> BMW Group, Knorrstr. 147, 80937 Munich, Germany

<sup>5</sup> Chair for Ergonomics, Technical University Munich, Boltzmannstr. 15, 85748 Garching, Germany; viktorija.geisel@tum.de

\* Correspondence: yannick.forster@bmw.de

† These authors contributed equally to this work.

Received: 13 November 2020; Accepted: 10 December 2020; Published: 14 December 2020



**Abstract:** During the last decade, research has brought forth a large amount of studies that investigated driving automation from a human factor perspective. Due to the multitude of possibilities for the study design with regard to the investigated constructs, data collection methods, and evaluated parameters, at present, the pool of findings is heterogeneous and nontransparent. This literature review applied a structured approach, where five reviewers investigated  $n = 161$  scientific papers of relevant journals and conferences focusing on driving automation between 2010 and 2018. The aim was to present an overview of the status quo of existing methodological approaches and investigated constructs to help scientists in conducting research with established methods and advanced study setups. Results show that most studies focused on safety aspects, followed by trust and acceptance, which were mainly collected through self-report measures. Driving/Take-Over performance also marked a significant portion of the published papers; however, a wide range of different parameters were investigated by researchers. Based on our insights, we propose a set of recommendations for future studies. Amongst others, this includes validation of existing results on real roads, studying long-term effects on trust and acceptance (and of course other constructs), or triangulation of self-reported and behavioral data. We furthermore emphasize the need to establish a standardized set of parameters for recurring use cases to increase comparability. To assure a holistic contemplation of automated driving, we moreover encourage researchers to investigate other constructs that go beyond safety.

**Keywords:** automated driving; literature review; human–automation interaction; method development

## 1. Introduction

The advent of automated driving (AD) systems and Human–Machine Interfaces (HMI) marks one of the biggest game changers in transportation research and development of our time. In 2013, the Society of Automotive Engineers [1] published the first version of their definition describing different levels of driving automation, which addresses challenges, sets foundation for future standardization, and establishes a common language. With SAE Level 2 (L2) driving automation

already on the road, it is only a matter of time until SAE Level 3 (L3) AD systems (ADS), or even higher levels, become commercially available. Thereby, the enormous potential of technological progress has sparked enthusiasm among human factors and Human–Computer Interaction researchers to develop user interfaces for this novel technology. Investigations of such interfaces through user studies are then conducted to first determine feasibility, and, in the next step, to fine-tune conceptual approaches. Here, basic research findings from engineering psychology, and human factors on one side, as well as computer science on the other, have been applied by automotive industry and academia. Furthermore, some lessons learned from prior automation development in the aviation sector [2–4] could be transferred. However, automated driving systems imply different preconditions which makes the situation much more complex. First, the driving environment is highly time-critical, and thus interventions must happen in seconds or even fractions of a second, while, in airplanes, pilots usually have more time to respond to critical events. Second, there exists greater variety among the targeted user groups. Automated vehicles are consumer products, and driver–passengers will have different levels of training and technology experience, acceptance of- and trust in automation, etc., and further come with additional goals (such as “driving fun” or the desire to engage in non-driving related activities) that are not that relevant, or not even present in classical operator settings. As a result, there are a wide range of challenges that need to be overcome, and a multitude of papers addressing these timely issues have been published over the last years. However, it is often hard to integrate and/or compare the obtained findings, as topics are often investigated differently. Furthermore, due to the sheer amount of results, it is hard for researchers to identify gaps where they could build upon.

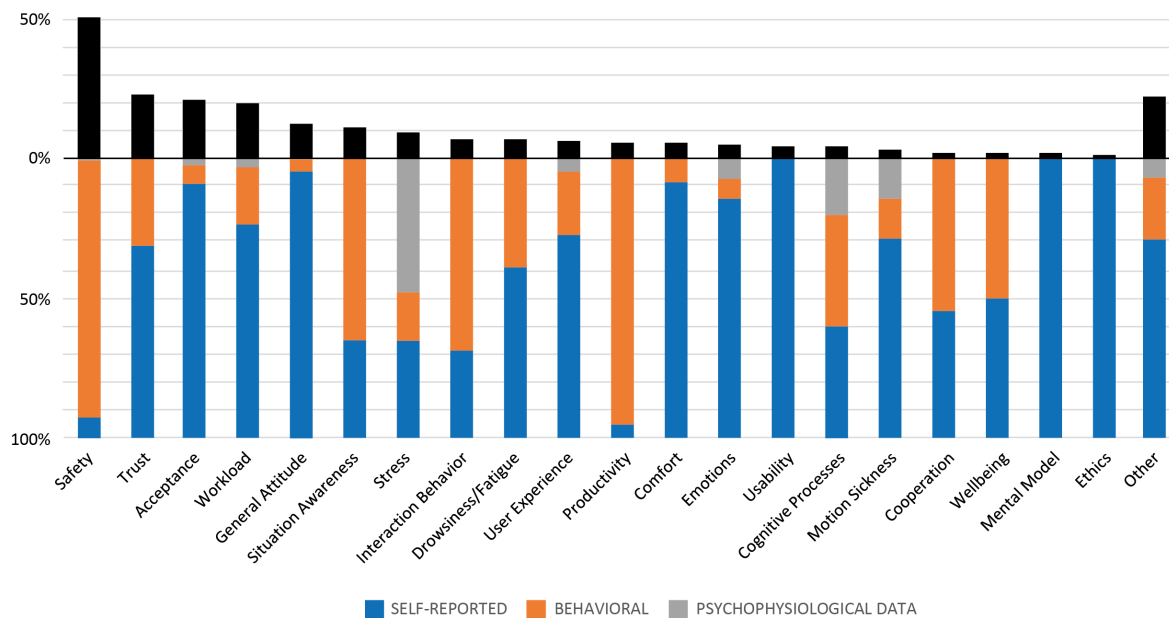
Thus, we claim that the time has come to systematically review which topics in driving automation received the most attention, and which methods have been applied to investigate these. Therefore, the present work aims at combining these approaches, and provide researchers and practitioners with an overview of current and past topics of AD. It also points towards improvements in the future, and unveils directions that have yet to be investigated. Moreover, we plan to give an impetus towards a standardization of methodology concerning self-report, behavioral, and physiological measures, as well as appropriate triangulation approaches of them. The main contribution of this work is twofold, concerning the status quo and derivation of future research directions:

- Providing an overview of emerging possibilities for study design in driving automation research
- Outlining which constructs have been investigated, which data collection methods have been applied, and which specific parameters have been calculated and reported

To reach this aim, we developed and followed a structured approach for reviewing related literature. First, we summarize important topics in driving automation that have been addressed in the last years, followed by a precise description of inclusion/exclusion criteria for publications used in this literature review. We then outline a procedure creating a database in which relevant contributions can be stored. Eventually, descriptive results of queries on the database are presented. The results of this literature review are expected to add to a better understanding of current trends and research directions of AD. Hence, it holds a mirror to this community on what has been accomplished and which future aspects need more attention (see Figure 1).

It is important to note here that the present paper does explicitly not target one specific area of research in automated driving. On the contrary, it serves the purpose of getting an overview of constructs on a higher level. From this overview, the work provides an in-depth summary of methodological approaches and measures within each of the constructs. This combination of high level comparison between research approaches (and their relative importance until now) as well as the insights into methodological approaches within each marks the novel contribution of this work. We acknowledge the fact that there is a large variety of research questions combined here. However, this work marks a first step to unveil trends in research on automated driving research. A more constrained work that focuses more specifically on one particular construct might be future research but is deliberately not part of the present work. Such work can also take into account the variety of research questions. Eventually, this work aims to come up with recommendations for

empirical user studies on automated driving. From the reasoning above, we are already aware that each recommendation might not be suitable for largely different research questions. However, the aim is to derive recommendations in a general manner so that researchers can adapt these to the specific research purpose. Moreover, these recommendations are not mandatory in a sense that we mean them to be forced upon researchers. In our opinion, the recommendations from such a profound database are meant for researchers to include these in their initial contemplation of study design.



**Figure 1.** Frequencies of studied constructs (upper row) and used data types (bottom row).

## 2. Research Status on Automated Vehicle HMIs

The HMI on the one hand shows information of the task (i.e., driving) to the user and on the other hand offers a possibility to provide input from the user to the driving automation system. User studies on automated vehicle (AV) HMIs have focused on different constructs, applied different collection methods, and also evaluated different parameters. In addition, different study design approaches are possible, which in turn depend on the respective construct and collection measure. However, there is no common agreed-upon methodological framework for evaluating AVs (and potential in-vehicle HMIs). Consortia reports are possible sources that researchers could consult [5–7]. Moreover, there have been first efforts to give an overview of methodological approaches in human–computer interaction research on user experience in general [8], the evaluation of in-vehicle information systems [9], on the evolution from manual to automated driving [10], and also with a focus on AD [11], however, without reviewing a broader set of categories.

### 2.1. Topics of Interest

Early research efforts in AD have often focused on Take-Over Request (TOR) scenarios, where the system exceeds its operational design domain or encounters an emergency, and prompts the driver to interrupt his/her non-driving related task (NDRT) to regain manual vehicle control. These transitions are either due to sensor failures/malfunctions (imminent transition) or because the system issues a planned indication to take over—in both cases, it is the user’s role to ensure a safe transition to manual driving. These studies have revealed different issues such as controllability [12–14], fatigue [15,16], mode awareness [17], or automation trust [18–22]. Other studies frequently applied survey approaches to investigate public acceptance and readiness for the introduction of this technology to the consumer market [23–25]. The downside of these acceptance-related studies is that, mostly, no realistic AD system is provided to the users. At best, a description of such a system is given that

requires a lot of imaginative power. A closely linked construct to acceptance is trust in automation. Here, there have been driving simulator studies that supported realistic ADS representations and HMIs [18,20,26–28]. Other topics that have recently emerged go beyond safety-related issues, such as usability [29,30], and user experience (UX) [31–33] of AD systems and HMIs. In the usability domain, research questions mainly focus on the measures and appropriate conditions under which users effectively and efficiently interact with driving automation and in turn are satisfied with the interface [34]. This also marks an additional factor in common acceptance frameworks [35,36]. User Experience [37] expands usability going beyond mere pragmatic aspects of using driving automation by adding hedonic qualities. With drivers being relieved of driving themselves, there might be a lack of fulfillment of needs [32,38–40], and, consequently, despite effectiveness and efficiency in interaction, positive emotions, thus, positive attitudes towards driving automation are not guaranteed. UX research therefore aims at investigating underlying mechanisms needed to substitute former driving experiences with other, potentially meaningful activities, and appropriate user interfaces which carve out advantages and balance negative effects of automation.

## 2.2. Possibilities for Study Design

Besides different topics of interest in AD HCI research, and thus constructs, collection methods, and parameters, which are all closely tied to dependent measures, there is also a variety of study design possibilities when conducting user studies on AD. One important aspect is the study environment which is a driving simulator in many instances. Here, the degree of immersion varies from low to medium fidelity simulation [41] to high fidelity driving simulation with [28,42] or without a moving-based platforms [13,43]. Moreover, depending on the availability of (real or simulated) automation function, studies on test tracks and real roads are possible [44,45]. Other types of studies use an interview or survey setting to gain insights into AD because the users have used such a system before [46,47], or they instead target the readiness of the consumer population [25,31]. Another aspect providing researchers with possibilities for study design is the representation of the automation. At the moment, most user studies in the AD context are set up in a simulation environment, since the maturity of this technology is not yet given. A vast majority of studies has been conducted in driving simulators, where AV functions and HMIs can be implemented without much effort, and tests in a risk-free and standardized environment are possible. However, first on-road tests have been conducted in Wizard-of-Oz [48,49], or even real settings [50,51]. As mentioned before, there are studies that are not placed in a lab environment, but rather take a survey approach and represent the driving automation by means of static descriptions [25], or sketches [31]. Studies also differ in the type of research or main focus and contribution they bear. Some studies target conceptual development with a subsequent proof-of-concept user study [52–54]. Other approaches more generically cover basic research topics, and focus on fundamentals of human perception and action in the AD context. Such research is rather independent from specific HMI concepts, but implications hold true for the variety of conceptual approaches [55,56]. A similar independence from certain HMI concepts is a characteristic of methodological work. Such studies aim at providing instruments and tools for proper study conduction when evaluating HMIs [29,30,57].

The heterogeneity of constructs, use cases, and AD operationalization has recently led to first efforts in methodological standardization. For example, various taxonomies [58–60] have been proposed to describe TOR scenarios and related use cases [61]. In addition, regarding TOR, many research groups follow their own measurement procedures and evaluation methods, which highlights the need for commonly accepted standards.

## 3. Method

With the considerations about the current research status, topics of interest in AD HCI research, study possibilities, and the authors' experience in AD research in mind, we defined the reviewing process for this literature review.

### 3.1. Venue Selection Process

We conducted a pilot study to identify journals and conferences in the Human Factors community that have published relevant and representative work on driving automation. An online survey was distributed via social media (e.g., Twitter, LinkedIn, etc.) and to peers of the authors. In the survey, participants could indicate (1) both the top 3 journals and conferences where they have already published as well as (2) both the top 3 journals and conferences where they consider submitting an article (favored). Moreover, the survey included questions on whether the authors have already published original research on automated driving (yes vs. no) and the year of the author's first publication. Eventually, demographic data (i.e., age, gender, academic degree, and academic background) were collected.

Demographics showed that mean age of the  $n = 21$  participants ( $n = 5$  female) was 32.81 years ( $SD = 5.65$ ,  $MIN = 24$ ,  $MAX = 49$ ). Most participants ( $n = 10$ ) held a Master's degree,  $n = 9$  a PhD, and  $n = 2$  were professors. The Academic Background showed that the majority were psychologists ( $n = 8$ ), followed by engineers ( $n = 5$ ), computer scientists ( $n = 5$ ).  $n = 2$  participants had a Human Factors or Media Informatics background (multiple choice was allowed). Out of the 21 participants,  $n = 19$  have already published whereas  $n = 2$  have not yet published their research. The earliest publication dates back to 1999 and the latest to 2016.

For the identification of relevant venues, we counted the overall number of instances independently from its position (i.e., 1st vs. 2nd vs. 3rd rank). The results regarding journals showed that Transportation Research Part F ( $n = 7$  publications,  $n = 11$  favored), the Journal of Human Factors ( $n = 7$  publications,  $n = 11$  favored) and Accident Analysis and Prevention ( $n = 7$  publications,  $n = 6$  favored) were the most frequently indicated venues. Regarding conferences, the AutomotiveUI ( $n = 6$  publications,  $n = 13$  favored), the Human Factors and Ergonomics Society Annual Meeting ( $n = 2$  publications,  $n = 11$  favored), and the Conference on Human Factors in Computing Systems (CHI;  $n = 2$  publications,  $n = 6$  favored) were mentioned most frequently. Based on this expert survey, we selected these three journals and conferences to be included in our structured literature review.

### 3.2. Paper Selection Process

The basis for the selection of papers for the present literature analysis were all research papers that were published in the respective venues between 2010 and up to 2018 (inclusive). We developed a decision tree to decide in a standardized and step-wise manner, whether or not to include each paper. This decision tree is depicted in Figure 2. It features four steps represented through binary decisions, where each has to be answered with "yes" for a paper to be included into our analysis. To pass the first step of the decision tree, the paper had to contain at least one of a set of keywords related to driving automation in the full text (see below). These keywords were selected to initially reduce the amount of papers in a reasonable way, while at the same time ensuring that no potentially relevant paper would be excluded. The first step of the decision tree was carried out by querying the respective online data bases using the following search terms:

*"automated driving" OR "autonomous driving" OR "self-driving" OR "self driving" OR "autonomous vehicle" OR "automated vehicle"*

Papers that did not feature AD represented by at least one of the keywords, as well as short papers, posters, and adjunct proceedings [62], were excluded in this step. For the remaining papers, the next step in the selection process consisted of examining whether the papers' objective was an empirical study to discard other literature reviews, as well as juridical, theoretical, or ethical papers [11,63,64]. The subsequent step of the decision tree aimed at the actual primary focus of the empirical papers. If this was not research and development of AD, the respective paper was excluded from further analysis. This step was incorporated to rule out work on concepts that, in principle, could be used for the development of AD, but was not originally investigated with that purpose [65]. In the last step of the selection process, we took a closer look at the levels of automation [1] that were investigated. The level of mere driver assistance (Level 1) as well as concepts which do not count as



driving automation according to SAE definition [66,67] were out of scope for this literature analysis. Thus, only papers examining L2 and/or higher levels of automation (i.e., simultaneous lateral and longitudinal vehicle control) were included.

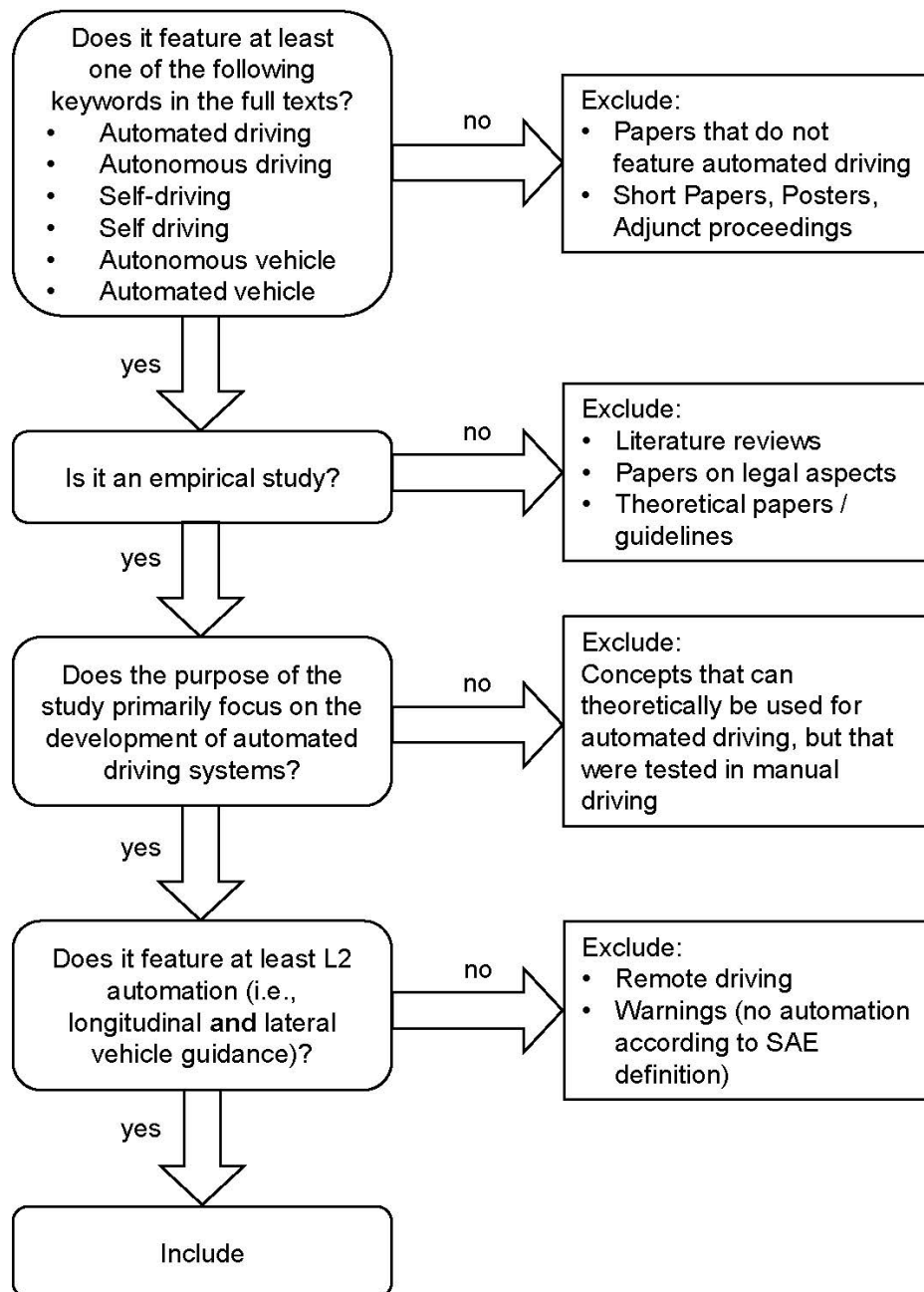


Figure 2. Decision tree for paper selection.

Overall,  $n = 161$  research papers passed all steps of the decision tree and were included in the present work. To ensure inter-rater reliability, a random selection of 10 papers were compared by means of intra-class correlations (ICC). ICC estimates were calculated based on a single-rating regarding the inclusion criteria (outcome of the decision tree to find out whether a paper should be included for further analysis), using a two-way random-effects model. Results revealed a high inter-rater reliability with a correlation of  $r = 0.809$  ( $F(9,36) = 22.170, p < 0.001$ ).

### 3.3. Paper Reviewing Strategy

After selecting the papers, we developed a reviewing strategy for literature analysis in two expert workshops. The first workshop lasted six hours and aimed at developing a standardized reviewing procedure. After the workshop, the resulting categories/dimensions as well as their emerging relations were translated into a database. For a detailed description of the categories and the database, see *Database Structure*. Subsequently, five reviewers classified the selected 161 research papers by sorting them into the categories of the database. In case a new category occurred in the papers that had not been considered before, it was added to the database. After reviewing a subset of the 161 papers, we conducted another expert workshop which lasted approximately four hours. During this workshop, lessons learned from a first subset of publications ( $n = 42$ ) were derived and each reviewer could put unclear classification up for discussion. This ensures potentially high reviewer agreement in classification of the remaining papers (ambiguities in the subset have been resolved during the workshop).

### 3.4. Database Structure

We set up a MS Access Database to capture the relevant information needed for our investigations. The schema consists of the five main tables *Paper*, *Conference*, *Construct*, *CollectionMethod*, and *Parameter*. In the following, we introduce the most relevant properties:

- **Paper:** For each paper, we collected descriptive information (title, abstract, year, authors, conference), as well as the levels of automation addressed in the study. The following additional information was collected: The type of user (driver, passenger, external), road type (urban, highway, rural, not relevant), study type (lab, test track, real road, survey), the representation of the AV (static text description, sketch, driving simulator, Wizard-of-Oz, real vehicle), study period (single session, short-term, long-term), type of research (basic research, concept evaluation, method development, model development), as well as participant information, such as the number of subjects, their mean age, as well as if they were internally (students, employees, etc.) or externally recruited.
- **Construct:** Represents the topics of investigation. To avoid subjective interpretation by the reviewers, we only collected constructs which were explicitly mentioned by authors in the papers, (such as *Safety*, *Trust*, *Acceptance*, etc.). All constructs which were only investigated by one single paper are summarized within another construct. Generic investigations on participants' opinion and general perception without directly mentioning specific constructs are summarized with a *General Attitude* construct.
- **Collection Method:** Relevant data collection methods, such as *driving performance*, *TOR performance*, *secondary task performance*, *EKG*, *EEG*, *standardized questionnaire*, *interview*, etc. Secondary task performance in particular refers to a participant's performance in a task not related to driving (e.g., number of missed target stimuli on a tablet). Again, we came up with initial suggestions that were expanded in case a new item emerged during the reviewing process.
- **Parameter:** Parameters that were used in the different data collection methods, for example standard deviation of lateral position (SDLP) [68], response time (could be used to measure driving performance), gaze-off-road time, gaze standard deviation (eye-tracking), technology acceptance model [35], NASA-TLX [69] (examples for standardized questionnaires), etc.
- **Relationships:** To structure our data, we created a relationship table to represent (n:m) relations of papers, collection methods, parameters, and constructs. Thus, each paper can investigate different constructs, where each construct can be assessed by one (or multiple) data collection methods, and each data collection method by one (or multiple) parameters. For each relation, we categorized whether the combination represents behavioral, self-reported, or physiological data. Furthermore, we assessed if the parameter was measured before/during/after a trial in the experiment. This data model allowed us to store all information without duplicates (each combination of paper/construct/collection method/parameter was stored only once using

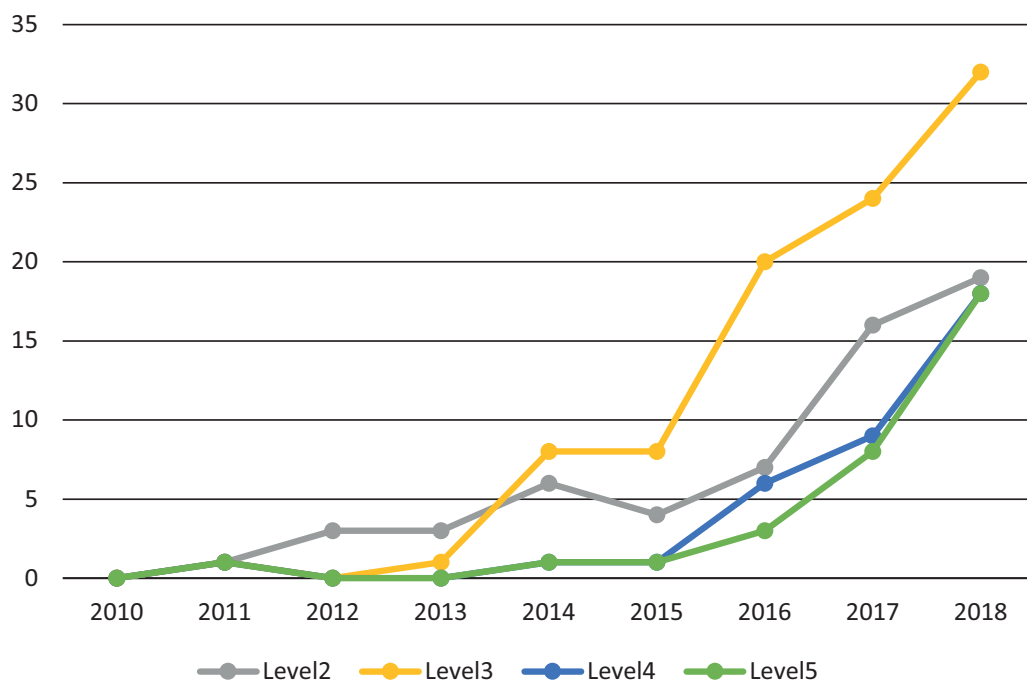
database key constraints), while the relations allow for performing powerful queries on the data (in comparison to pure list/sheet based representations). For example, we can ask the database which papers investigated a certain construct using physiological sensors, how many papers utilized a certain standardized scale, or how constructs changed over the years.

#### 4. Results

In the following, we report the obtained results from of the final selection of 161 papers. All selected conferences and journals (see above) are represented in the final analysis: Human Factors and Ergonomics Society Annual Meeting (n = 34), AutomotiveUI (n = 42), CHI (n = 10), Accident Analysis (n = 18), Human Factors (n = 20), and Transportation Research Part F (n = 37). All results were obtained using the built-in structured query language (SQL) of MS-Access.

##### 4.1. General Study Details

Regarding AD research, we found that SAE L3 is the most frequently studied level of automation with 58.39% (n = 94) followed by Level 2 (36.65%, n = 59) and Level 4 (22.36%, n = 36). Level 5 was investigated in 19.88% (n = 32) of the studies. Thereby, the number of publications gradually increased up to 2018 regardless of a specific level of automation. The one that stands out is SAE L3, which attracted earlier attention than the other levels (see Figure 3); however, the steepest increase in AD research was observed between 2015 and 2016.



**Figure 3.** Frequencies (n) of ADS papers regarding the investigated SAE level of automation as evolved over time since 2010. SAE L0 and L1 are excluded as they are not part of this literature review.

Overall, 73.29% (n = 118) of all studies were conducted in a lab environment, 13.66% (n = 22) as a survey, and 11.80% (n = 19) on real road, while only 2.48% (n = 4) reported results obtained on a test track. This is in accordance with the utilized AV representation. In addition, 71.43% (n = 115) of the papers reported to have used a driving simulator, 12.42% (n = 20) a real vehicle, 11.18% (n = 18) a textual description, 7.45% (n = 12) a Wizard-of-Oz setup, and 3.11% (n = 5) a visualization. We observed thereby a clear tendency towards single session evaluation (94.41%, n = 152), and only nine papers called in participants multiple times, where 3.11% (n = 5) investigated automated driving use over a

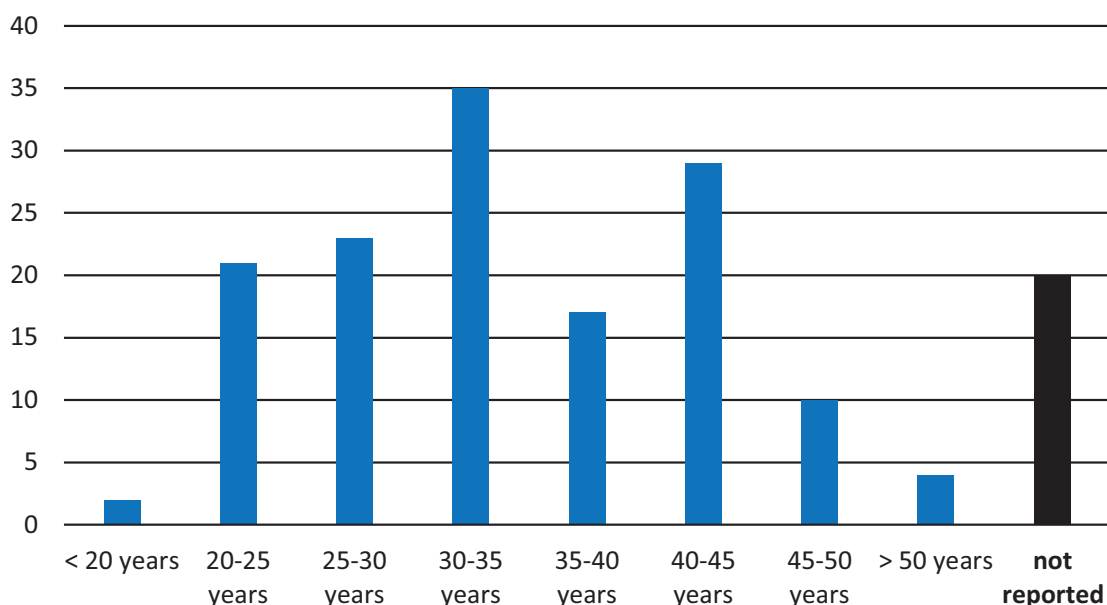


short time period (e.g., up to one week) [70] and three studies (1.86%) long-term effects in longitudinal studies (e.g., by following a survey approach [47]).

While 60.25% (n = 97) of the papers conducted basic research such as observing pedestrian interaction behavior with AVs on real roads [71], 34.78% (n = 56) papers evaluated a specific concept in their study (e.g., a haptic seat to prepare driver for TORs) [72]. Smaller percentages of studies conducted empirical research about AD with the aim to create a method (8.07%, n = 13) or a model (1.86%, n = 3). The center of empirical research on AD is clearly the driver, who was investigated in 78.26% (n = 126) of all studies. Passengers (9.32%, n = 15) and other road users (6.83%, n = 11, e.g., pedestrians) are still a side topic in AD HCI research. The environmental driving context is varying, while 47.83% (n = 77) investigated a highway setting, 19.88% (n = 32) addressed urban, and 19.25% (n = 31) rural road conditions. For the remaining studies (13.04%, n = 21), the road type was either not relevant or not described.

Study participants' mean age in the most papers is below 40 years (n = 98), and only a few papers report higher mean ages such as Frison et al. [73], who in particular invited participants older than 65 to investigate acceptance of AVs in comparison with participants of younger age (see Figure 4). In 12.42% (n = 20), participants' age is not reported.

In total, 57.76% (n = 93) of the papers selected their participants randomly, while only 29.81% (n = 48) used targeted sampling. The rest (n = 20) did not provide any indication about participant sampling. Thereby, 34.40% (n = 57) of papers recruited participants internally (e.g., students or employees of their institution), and 44.10% (n = 71) invited external participants. The remaining publications (20.5%) did not provide this information.



**Figure 4.** Frequencies (n) of age groups as used in the analyzed automated driving studies.

#### 4.2. Methodological Approaches

Regarding methodological approaches, we observed that 74.91% (n = 119) of all papers involve self-reported data collection, 64.60% (n = 104) behavioral, and 10.56% (n = 17) psycho-physiological data.

Over half of the papers triangulated different types of data (55.28%, n = 89). Thereby, 44.72% (n = 72) of the papers triangulated behavioral and self-reported data. Less common is the triangulation of self-reported, behavioral and psycho-physiological data (6.21%, n = 10). The combination of behavioral and psycho-physiological data (n = 5), or self-reported and psycho-physiological data (n = 2) is even more rarely applied. A large portion of the papers (44.72%, n = 72) is working with only

one type of data. These papers use mainly self-report measures (27.95%, n = 45), while 16.77% (n = 27) of the papers report exclusively behavioral data.

Overall, we identified n = 22 different collection methods. The self-defined questionnaire is the most frequently used method (51.85%, n = 84) in AD research papers. This is in accordance with the large number of self-reported data collection. Standardized questionnaires (44.44%, n = 72) and TOR performance (37.65%, n = 61) are frequently used as well. Eye tracking/gaze behavior (24.07%, n = 39), driving performance measures (22.22%, n = 36), interviews (16.05%, n = 26) and secondary task/NDRT performance (8.64%, n = 14) are used more seldomly. Furthermore, rarely applied collection methods are special techniques such as heart rate variability (3.70%, n = 6), think aloud (2.47%, n = 4), probing, electroencephalography (EEG) and detection tasks (each 1.85%, n = 3), UX curve, sorting, electromyography (EMG), and galvanic skin response (GSR; each 1.23%, n = 2). Methods only used once are near infrared spectroscopy (NIRS), focus group, facial expression detection, and matching tasks. Details of the applied methods are described in the following section.

### 4.3. Constructs and Methodological Approaches

To specify empirical research on AD in more detail, we took a closer look at the constructs, which were investigated in the individual studies. The most frequently investigated construct is *safety*, followed by *trust*, *acceptance*, and *workload*, see Table 1 for all constructs, respective number of papers and exemplary references. We observed that there was a broad range of 36 distinct constructs that was addressed only once. We summarized it here as *other*.

**Table 1.** Constructs investigated by papers, with a non-exhaustive list of example papers for each category.

	n	in %	Examples
Safety	82	50.93	Gold et al. [13], Merat et al. [14], Wintersberger et al. [21]
Trust	37	22.98	Hergeth et al. [19], Forster et al. [74]
Acceptance	34	21.12	Nordhoff et al. [23], Kyriakidis et al. [25]
Workload	32	19.88	Bashiri and Mann [75], Biondi et al. [76]
General Attitude	20	12.42	Maurer et al. [77]
Situation Awareness	18	11.18	Clark et al. [78], Karjanto et al. [79]
Stress	15	9.32	Wintersberger et al. [21]
Interaction Behavior	11	6.83	Currano et al. [71]
Drowsiness/Fatigue	11	6.83	Neubauer et al. [16], Naujoks et al. [80]
User Experience	10	6.21	Frison et al. [32,73], Oliveira et al. [81]
Productivity	9	5.59	Chang et al. [54], Reimer et al. [82]
Comfort	9	5.59	Bellem et al. [83]
Emotions	8	4.97	Wintersberger et al. [26], Terken et al. [84]
Usability	7	4.35	Forster et al. [41]
Cognitive Processes	7	4.35	Glatz et al. [85]
Motion Sickness	5	3.11	Karjanto et al. [79]
Cooperation	3	1.86	Guo et al. [86], Walch et al. [87]
Wellbeing	3	1.86	Telpaz et al. [72]
Mental Model	3	1.86	Beggiato et al. [88]
Ethics	2	1.24	Sikkenk and Terken [56]
Other, e.g., Personalization	36	22.36	-

In the following paragraphs, individual constructs, summarized in the subsection according their occurrence of investigation, are described by elaborating on the applied collection methods (n ≙ number of distinct papers) and collected parameters (n<sub>p</sub> ≙ total number of parameters). One distinct paper can investigate more than one parameter.

### 4.3.1. Safety

Parameters (p) selected for the AD studies on *safety* mainly include behavioral data (only  $n_p = 22$  collected self-reported data;  $n_p$  represents the number of parameters). The most applied collection method for *safety* is the measurement of participants’ TOR performance, which is applied by  $n = 58$  distinct papers. Thereby, the most collected parameter is participants’ reaction time, which includes the time to first driving action like braking or accelerating, to system deactivation, button press or hands on the steering wheel. Furthermore, the lateral position is another frequently utilized parameter, including maximal lateral position, standard deviation of lateral position, or Daimler Lane Change Performance [89]. Furthermore, Time to Collision (TTC), speed TOR timings and acceleration and braking parameters were also repeatedly collected. Driving performance (which is in contrast to TOR performance that assesses the immediate response, calculated based on longer phases of manual driving) is also often used ( $n = 24$ ) by e.g., collecting data on participants lateral position, speed and reaction times. Eye-tracking ( $n = 12$ ) by regarding gaze percentage, duration and number on areas of interests like mirrors or road, etc., observation ( $n = 12$ ) of participants crossing behavior, NDRT engagement parameters, etc., and a self-defined questionnaire ( $n = 11$ ) are further important collection methods for safety. Standardized questionnaires like the scale for criticality assessment of driving and traffic scenarios [90], secondary task performance and interviews are more seldom used ( $n \leq 3$ ), see Table 2 for more details.

In total,  $n_p = 89$  different parameters were collected across all papers investigating safety—83% of the studies were conducted in a driving simulator, 11% in a real vehicle, 5% used static text, 4% a Wizard-of-Oz setup, and 2% a sketch.

**Table 2.** Safety Collection methods and parameters.

Collection Method for Safety (n)	Parameter	$n_p$
TOR Performance (58)	Reaction Time	70
	Lateral Position	18
	Time to Collision	15
	Speed Parameters	10
	TOR Timing	10
	Acceleration	9
	Braking	7
	Steering Wheel Angle	5
	Distance Front	4
	First, Driving Action	4
	Number of Collisions	4
	Disengagements	3
	Lane Change Parameters	3
	Steering	3
	Accuracy	2
	NDRT Engagement	2
	Accident Avoidance Ranking (AAR)	1
	Number of Interactions	1
N/D	7	

Table 2. Cont.

Collection Method for Safety (n)	Parameter	n <sub>p</sub>
Driving Performance (24)	Lateral Position	13
	Speed Parameters	9
	Reaction Time	7
	Time to Collision	5
	Acceleration	3
	Steering Wheel Angle	3
	Braking	2
	Lane Departure Parameters	2
	Number of Collisions	2
	Automation Enabled/Disabled	1
	Distance Front	1
	Overtakings per Km	1
	Steering	1
	N/D	1
Eye Tracking/Gaze Behavior (12)	Gaze Percentage	9
	Gaze Duration	6
	Gaze Number	3
	Glancing Behavior	3
	Reaction Time	3
	Pathways	1
	Saccade	1
Observation (12)	Crossing Behavior	2
	NDRT Engagement	2
	Reaction Time	2
	Accuracy	1
	Automation Enabled/Disabled	1
	Braking	1
	Crossing Time	1
	Gaze Number	1
	Number of Collisions	1
	Situation Criticality	1
	Steering	1
	Time to Collision	1
	N/D	2
Self-Defined Questionnaire (11)	N/D	10
	Accuracy	1

Table 2. Cont.

Collection Method for Safety (n)	Parameter	n <sub>p</sub>
Standardized Questionnaire (6)	Scale for Criticality Assessment of Driving and Traffic Scenarios	2
	Cooper–Harper Scale	2
	Auditory Urgency Scale	1
Secondary Task Performance (3)	NDRT Engagement	1
	Reaction Time	1
	N/D	1
Interviews (2)	Semi-structured Interview	2
Matching (1)	Accuracy	1

#### 4.3.2. Trust, Acceptance, and Workload

*Trust* in AD is investigated as the second most used construct, see Table 1. Here, in contrast to safety, more self-reported ( $n_p = 42$ ) data are reported. Most common is the usage of a self-defined questionnaire ( $n = 19$ ), or a standardized questionnaire, especially the Automation Trust Scale (ATS) [91] ( $n = 12$ ) is popular. The interpersonal trust scale (ITS) [92], trust in technology scale [93], Van der Laan scale [94], the Trust Perception Scale-HRI [95] and the Propensity to Trust Scale [96] are each used only once. Interviews, structured and semi-structured, are more rarely used ( $n = 4$ ). However, behavioral parameters ( $n_p = 19$ ) are also collected, by observing ( $n = 5$ ) body pose and movements, acceleration and braking behavior, gaze duration on an area of interest, etc. In addition, eye-tracking ( $n = 4$ ) is conducted. Only one paper [97] used driving performance (braking and steering behavior) and TOR (reaction time) performance measures. This paper additionally collected participants' hands on wheel and eyes on road time using observation, see Table 3. In total,  $n_p = 28$  different parameters were collected across all papers investigating trust—78% of the studies were conducted in a driving simulator, 16% in a real vehicle, 11% used a textual description and 5% a Wizard-of-Oz setup to represent the AV.

*Acceptance* is investigated by self-reported parameters as intensively as *trust* ( $n_p = 41$ ). Researchers mostly apply standardized questionnaires ( $n = 18$ ). Thereby, the Van der Laan acceptance scale [94] is the most frequently used questionnaire. Furthermore, the technology acceptance model (TAM) [35], the Unified Theory of Acceptance and Use of Technology (UTAUT) [36] or the Car Technology Acceptance Model (CTAM) [24] are frequently used approaches. The self-defined questionnaire is also a popular collection method here ( $n = 13$ ). Behavioral data ( $n_p = 3$ ) on acceptance is collected by observing ( $n = 3$ ) the time to system activation, and the number/share of times the automation was enabled/disabled by study participants. Driving performance, or qualitative methods like interviews or focus groups, are applied less frequently ( $n < = 2$ ), see Table 4.

In total,  $n_p = 15$  different parameters were collected across all papers investigating acceptance. In total, 53% of the studies were conducted in a driving simulator, 29% used only static text, and 15% used a real vehicle, respectively 3% used either a sketch or a Wizard-of-Oz setup.

**Table 3.** Trust collection methods and parameters.

Collection Method for Trust (n)	Parameter	n <sub>p</sub>
Self-Defined Questionnaire (19)	N/D	19
Standardized Questionnaire (17)	Automation Trust Scale (ATS)	12
	Interpersonal Trust Scale (ITS)	1
	Van der Laan Acceptance Scale	1
	Trust in Technology Scale	1
	Trust Perception Scale-HRI	1
	Propensity to Trust Scale	1
	N/D	1
Observation (5)	Body Pose/Movements	4
	Acceleration	1
	Brake	1
	Driving Action	1
	Gaze Duration	1
	Reaction Time	1
	Steering	1
Interviews (4)	Semi-structured Interview	3
	Structured Interview	1
Eye Tracking/Gaze Behavior (4)	Gaze Duration	2
	Gaze Percentage	1
	Glancing Behavior	1
Driving Performance (1)	Gaze Number	1
	Brake	1
Decision Game (1)	Steering	1
	N/D	1
TOR Performance (1)	Reaction Time	1

Workload is also investigated mainly by using self-reported measures (n<sub>p</sub> = 26), usually by implementing standardized questionnaires. The NASA Task Load Index (NASA-TLX, n = 17) [69] is the most popular, other questionnaires like Driver Activity Load Index (DALI) [98], Rating Scale Mental Effort (RSME) [99], scale for subjectively experienced effort (SEA) [100], or the global mental workload measurement by Wierwille and Casali [101] are more rarely applied, as well as self-defined questionnaires (n = 4) and semi-structured interviews (n = 1). As behavioral data (n<sub>p</sub> = 7), the secondary task performance collecting data on NDRT performance (i.e., number solved tasks), using the Surrogate Reference Task, or Twenty Question Task, observation, eye-tracking and driving performance measures were conducted, see Table 5.

In total, n<sub>p</sub> = 12 different parameters were collected across all papers investigating workload—88% investigated it in a driving simulator, only 13% in a real vehicle, and 6% used a Wizard-of-Oz setup.



**Table 4.** Acceptance collection methods and parameters.

Collection Method for Acceptance	Parameter	n <sub>p</sub>
Standardized Questionnaire (18)	Van der Laan Acceptance Scale	10
	Unified Theory of Acceptance and Use of Technology (UTAUT)	4
	Technology Acceptance Model (TAM)	3
	Car Technology Acceptance (CTAM)	2
	Willingness to Ride	1
	Perceived Behavioral Control (PCB)	1
	System Usability Scale	1
	Personal Innovativeness Scale	1
Self-Defined Questionnaire (13)	N/D	13
Observation (3)	Automation Enabled/Disabled	1
	Reaction Time	1
	N/D	1
Driving Performance (1)	Proportion of Manually Driven Scenarios	1
Interviews (1)	Unstructured Interview	1
Focus Group (1)	N/D	1
Secondary Task Performance (1)	NDRT engagement	1

**Table 5.** Workload collection methods and parameters.

Collection Method for Workload (n)	Parameter	n <sub>p</sub>
Standardized Questionnaire (22)	NASA-TLX	17
	Driver Activity Load Index (DALI)	2
	Rating Scale Mental Effort (RSME)	1
	Scale for Subjectively Experienced Effort (SEA)	1
	Global Mental Workload Measurement	1
Self-Defined Questionnaire (4)	N/D	4
Secondary Task Performance (4)	NDRT Performance	2
	Twenty Question Task (TQT)	1
	Surrogate Reference Task (SURT)	1
Observation (1)	N/D	1
Interviews (1)	Semi-Structured Interview	1
Eye Tracking/Gaze Behavior (1)	Glancing Behavior	1
Driving Performance (1)	Steering Wheel Angle	1

#### 4.3.3. General Attitude, Situation Awareness, and Stress

Participants' *general attitude* towards AD is investigated by the majority of the papers by self-reported data (n<sub>p</sub> = 22), by a self-developed questionnaire (n = 15) or interviews (n = 5). One paper derived insight based on observation (see Table 6).

In total,  $n_p = 3$  different parameters were collected across all papers investigating general attitude—40% used a driving simulator, 30% static text, 15% a real vehicle, and 15% a Wizard-of-Oz setup.

**Table 6.** General Attitude collection methods and parameters.

Collection Method for General Attitude (n)	Parameter	$n_p$
Self-Defined Questionnaire (15)	N/D	15
Interviews (5)	Semi-Structured Interview	4
	Unstructured Interview	1
Standardized Questionnaire (2)	BIG 5	2
Observation (1)	N/D	1

*Situation Awareness* in contrast is examined more by behavioral ( $n_p = 26$ ) than self-reported data ( $n_p = 14$ ). However, almost the same share of papers apply a self-developed questionnaire ( $n = 6$ ) as eye-tracking is used ( $n = 7$ ). Thereby, gaze duration, number, and percentage is collected, as well as glancing and blinking behavior. In addition, TOR performance ( $n = 2$ ) measures were collected like lateral position, reaction time, acceleration, and time to collision parameters. A popular standardized questionnaire is the situational awareness rating technique (SART), or the situation awareness global assessment technique (SAGAT) [102]. See Table 7. In total,  $n_p = 25$  different parameters were collected across all papers investigating situation awareness—67% of the studies were conducted used a driving simulator, 28% a real vehicle, and 17% a Wizard-of-Oz setup was used.

**Table 7.** Situation Awareness collection methods and parameters.

Collection Method for Situation Awareness (n)	Parameter	$n_p$
Eye Tracking/Gaze Behavior (7)	Gaze Duration	5
	Gaze Number	4
	Gaze Percentages	3
	Glancing Behavior	2
	Blink Behavior	1
	Reaction Time	1
	Self-Defined Questionnaire (6)	N/D
Standardized Questionnaire (3)	Situational Awareness Rating Technique (SART)	3
Probing (2)	Situation Awareness Global Assessment Technique (SAGAT)	3
TOR Performance (2)	Lateral Position	2
	Reaction Time	2
	Acceleration	1
	Time to Collision	1
Interviews (2)	Semi-structured Interview	2
Observation (1)	Accuracy	1
	Reaction Time	1
Secondary Task Performance (1)	NDRT Performance	1

Stress is investigated by psycho-physiological ( $n_p = 11$ ) and self-rated data ( $n_p = 8$ ). Hence, standardized questionnaires like the Short Stress State Questionnaire (SSSQ) [103], Dundee Stress State Questionnaire (DSSQ) [104], or the Driver Stress Inventory (DSI) [105] are used, but also heart rate variability, GSR and EMG. Observation, interviews and driving performance measures were only used by single papers. See Table 8. In total,  $n_p = 17$  different parameters were collected across all papers investigating stress—80% of all studies regarding stress were performed in a driving simulator and 20% in a real vehicle—respectively, 7% used either static text or a Wizard-of-Oz setup.

**Table 8.** Stress collection methods and parameters.

Collection Method for Stress (n)	Parameter	$n_p$
Standardized Questionnaire (6)	Short Stress State Questionnaire (SSSQ)	3
	Dundee Stress State Questionnaire (DSSQ)	2
	Driver Stress Inventory (DSI)	1
Heart Rate Variability (4)	HR (BPM)	2
	Physical Position	1
	Root Mean Square of Successive Differences (RMSSD)	1
GSR (2)	AmpSum	1
	ISCR	1
	nSCR	1
	SCR	1
	PhasicMax	1
	N/D	1
Eye Tracking/Gaze Behavior (1)	Gaze Duration	1
	Gaze Number	1
Self-Defined Questionnaire (1)	Other	1
Observation (1)	Body Pose/m Movements	1
Interviews (1)	Semi-structured Interview	1
EMG (1)	N/D	1
Driving Performance (1)	Automation Enabled/Disabled	1

#### 4.3.4. Interaction Behavior, Drowsiness/Fatigue, and User Experience

Participants' *interaction behavior* with AV/ADS was mainly investigated by behavioral data ( $n_p = 11$ ), observing ( $n = 6$ ) such as pedestrians' walking behavior, number/share of times the automation was activated/deactivated, or eye-tracking analysis (percentage of the gaze on different AOIs). However, standardized questionnaires ( $n = 3$ ) like the Pedestrian Behavior Questionnaire (PBQ) [106], Brief Sensation Seeking Scale (BSSS-8) [107] or the theory of planned behavior (TPB) [108] were also used. Two papers developed an own questionnaire, and one paper applied eye-tracking, conducted an interview or collected data on secondary task performance, see Table 9. In total,  $n_p = 13$  different parameters were collected across all papers investigating interaction behavior—45% studies regarding interaction behavior were investigated in a driving simulator, 36% used a Wizard-of-Oz setup, and 27% a real vehicle.

**Table 9.** Interaction Behavior collection methods and parameters.

Collection Method for Interaction Behavior (n)	Parameter	n <sub>p</sub>
Observation (6)	Walking Behavior	2
	Automation Enabled/Disabled	2
	Glancing Behavior	1
	Pathways	1
	Reaction Time	1
	NDRT Engagement	1
	N/D	2
Standardized Questionnaire (3)	Brief Sensation Seeking Scale (BSSS-8)	1
	Pedestrian Behavior Questionnaire (PBQ)	1
	Theory of Planned Behavior (TPB)	1
Self-Defined Questionnaire (2)	N/D	2
Eye Tracking/Gaze Behavior (1)	Gaze Percentages	1
Interviews (1)	Semi-structured Interview	1
Secondary Task Performance (1)	Single-choice Quiz	1

*Drowsiness/Fatigue* was investigated by self-reported (n<sub>p</sub> = 11) and behavioral data (n<sub>p</sub> = 7). Hence, standardized questionnaires (n = 7) like Karolinska Sleepiness Scale (KSS) [109], Driver Stress Inventory (DSI) [110], Multidimensional Fatigue Inventory (MFI) [111] are mainly used as well as self-developed questionnaires (n = 2). However, researchers also collected (n = 2) blinking behavior or yawning, as well as eye-tracking, driving performance or adapted the method of UX Curve [112]. See Table 10. In total, n<sub>p</sub> = 12 different parameters were collected across all papers investigating drowsiness/fatigue—91% of the studies collecting drowsiness were conducted in a driving simulator, 9% used a real vehicle, and a further 9% used static text as AV representation.

**Table 10.** Drowsiness/Fatigue collection methods and parameters.

Collection Method for Drowsiness/Fatigue (n)	Parameter	n <sub>p</sub>
Standardized Questionnaire (7)	Karolinska Sleepiness Scale (KSS)	4
	Driver Stress Inventory (DSI)	1
	Multidimensional Fatigue Inventory (MFI)	1
	Self-Assessment Manikin (SAM) Scale	1
	Dundee Stress State Questionnaire (DSSQ)	1
Self-Defined Questionnaire (2)	Other	2
Observation (2)	Yawning	1
	Blink Behavior	1
	N/D	1
Driving Performance (1)	Reaction Time	2
	Lateral Position	1
Eye Tracking/Gaze Behavior (1)	Glancing Behavior	1
UX-Curve (1)	N/D	1

User Experience is mainly investigated by self-reported measurements  $n_p = 16$ ,  $n_p = 5$  collect behavioral data, only once psycho-physiological measures by capturing Heart Rate Variability is utilized. Various standardized questionnaires ( $n = 4$ ) like AttrakDiff [113], UEQ [114], etc. are applied, or interviews ( $n = 4$ ), and other qualitative methods like UX curve [112], think aloud, and sorting have been conducted. Behavioral data are collected by one paper regarding participants' driving performance, e.g., acceleration, braking, speed, and lane changes. See Table 11. In total,  $n_p = 13$  different parameters were collected across all papers investigating UX.

**Table 11.** UX collection methods and parameters.

Collection Method for UX	Parameter	$n_p$
Standardized Questionnaire (4)	AttrakDiff	2
	User Experience Questionnaire (UEQ)	1
	Van der Laan Acceptance Scale	1
	Hedonia and Eudaimonia (HEMA) Scale	1
	Sheldon's Need Scale	1
Interviews (4)	Semi-structured interview	4
Self-Defined Questionnaire (2)	<i>N/D</i>	2
Driving Performance (1)	Acceleration	1
	Brake	1
	Speed Parameters	1
	Number of Lane Changes	1
Observation (1)	<i>N/D</i>	3
Heart Rate Variability (1)	HR (BPM)	1
UX-Curve (1)	<i>N/D</i>	1
Think Aloud (1)	<i>N/D</i>	1
Sorting (1)	<i>N/D</i>	1

#### 4.3.5. Productivity, Comfort, Emotions, Usability, Cognitive Processes, and Motion Sickness

Researchers interested in *productivity* collect behavioral data  $n_p = 20$ , primarily about participants' secondary task performance ( $n = 6$ ), including performance (e.g., characters per second, number of answered questions, etc.) but also engagement and accuracy (e.g., error rate) parameters (NDRT duration, frequency, or percentage). Single papers also observed participants, used eye-tracking or collected driving performance measures. In addition, only one semi-structured interview was conducted, see Table 12. In total,  $n_p = 18$  different parameters were collected across all papers investigating productivity.

**Table 12.** Productivity collection methods and parameters.

Collection Method for Productivity (n)	Parameter	n
Secondary Task Performance (6)	NDRT Performance	8
	NDRT Engagement	3
	Accuracy	3
	Body Pose/Movements	1
	N/D	1
Driving Performance (1)	Accuracy	1
	Reaction Time	1
Observation (1)	Accuracy	1
Interviews (1)	Semi-structured Interview	1
Eye Tracking/Gaze Behavior (1)	Gaze Percentages	1

On the contrary, *comfort* was investigated instead by self-reported data ( $n_p = 11$ ), mainly by self-defined ( $n = 7$ ) or standardized questionnaires ( $n = 2$ ), Driving Style Questionnaire (MDSI), UEQ, TAM, and UTAUT. One paper reported behavioral data about participants' acceleration, see Table 13. In total,  $n_p = 6$  different parameters were collected across all papers investigating comfort.

**Table 13.** Comfort collection methods and parameters.

Collection Method for Comfort (n)	Parameter	$n_p$
Self-Defined Questionnaire (7)	N/D	7
	Driving Style Questionnaire (MDSI)	1
	Technology Acceptance Model (TAM)	1
Standardized Questionnaire (2)	Unified Theory of Acceptance and Use of Technology (UTAUT)	1
	User Experience Questionnaire (UEQ)	1
Driving Performance (1)	Acceleration	1

In addition, *emotions* are investigated by self-reported data ( $n_p = 12$ ) using self-defined ( $n = 4$ ) and standardized questionnaires ( $n = 4$ ) like PANAS, Affect Grid, Multi-Modal Stress Questionnaire (MMSQ), Affect Scale [115], etc. Furthermore, emotions are investigated by observing facial expressions, think aloud technique or an interview (each  $n = 1$ ). See Table 14. In total,  $n_p = 9$  different parameters were collected across all papers investigating emotions.

Moreover, *usability* is investigated as well mainly by self-developed ( $n = 4$ ) and standardized questionnaires ( $n = 3$ ), e.g., SUS is a popular method. Furthermore, semi-structured interviews ( $n = 2$ ) and the think aloud technique are also applied. Hence, solely self-reported data are collected ( $n_p = 11$ ), see Table 15. In total,  $n_p = 4$  different parameters were collected across all papers investigating usability.



**Table 14.** Emotions collection methods and parameters.

Collection Method for Emotions (n)	Parameter	n <sub>p</sub>
Standardized Questionnaire (4)	Positive and Negative Affect Scale (PANAS)	2
	PANAS-X	1
	Affect Grid	1
	Multi-Modal Stress Questionnaire (MMSQ)	1
	Affect Scale	1
Self-Defined Questionnaire (4)	N/D	3
	Russel's Circumplex Model	1
Observation (1)	Percentage of Detected Emotions	1
	Facial Expressions	1
Interviews (1)	Semi-structured Interview	1
Think Aloud (1)	N/D	1

**Table 15.** Usability collection methods and parameters.

Collection Method for Usability (n)	Parameter	n <sub>p</sub>
Self-Defined Questionnaire (4)	N/D	4
Standardized Questionnaire (3)	System Usability Scale (SUS)	3
	Input-Output Questionnaire	1
Interviews (2)	Semi-structured Interview	2
Think Aloud (1)	N/D	1

*Cognitive processes*, in contrast, are analyzed by self-reported ( $n_p = 4$ ), behavioral ( $n_p = 4$ ), and psycho-physiological data ( $n_p = 2$ ). Thereby, most developed their own questionnaires ( $n = 3$ ) and one paper used a standardized questionnaire, the Driver Stress Inventory (DSI). EEG is utilized twice, while respectively one paper applied driving performance measures (e.g., lateral position) or a detection task method collecting reaction time and accuracy. See Table 16. In total,  $n_p = 6$  different parameters were collected across all papers investigating cognitive processes.

**Table 16.** Cognitive Processes collection methods and parameters.

Collection Method for Cognitive Processes	Parameter	n <sub>p</sub>
Self-Defined Questionnaire (3)	N/D	3
EEG (2)	N/D	2
Detection Task (1)	Accuracy	1
	Reaction Time	1
Driving Performance (1)	Lateral Position	1
	Time Headway	1
Standardized Questionnaire (1)	Driver Stress Inventory (DSI)	1

*Motion Sickness* is mainly investigated by the standardized questionnaires ( $n = 4$ ), the Simulator Sickness Questionnaire (SSQ) [116], and the Motion Sickness Assessment Questionnaire (MSAQ) [117]. One paper developed an own questionnaire. However, Heart Rate Variability as well as the measure of Motion Sickness Dose Value were collected, see Table 17. In total,  $n_p = 5$  different parameters were collected across all papers investigating motion sickness.

**Table 17.** Motion Sickness collection methods and parameters.

Collection Method for Motion Sickness (n)	Parameter	n <sub>p</sub>
Standardized Questionnaire (4)	Sickness Questionnaire (SSQ)	3
	Motion Sickness Assessment Questionnaire (MSAQ)	1
Self-Defined Questionnaire (1)	Other	1
Heart Rate Variability (1)	HR (BPM)	1
Driving Performance (1)	Motion Sickness Dose Value (MSDV)	1

#### 4.3.6. Other Constructs

A wide range of additional more special constructs were investigated, including *cooperation*, *well-being*, *mental models* and *ethics* in which only a few papers were interested in (n < 3), and we identified many special constructs only single papers were interested in (n = 1), e.g., personalizing, intuitiveness, immersion, helpfulness, annoyance, motivation, etc. While *cooperation* was investigated mainly by driving performance measures like reaction time, duration of vehicle interaction, etc., and self-defined questionnaires and *well-being* by self-defined and standardized questionnaires and observation, researchers interested in *mental models* and *ethics* solely applied self-defined questionnaires. The more special constructs were summarized here as others were investigated in most cases by self-defined questionnaires (n = 15) or specific standardized questionnaires (n = 12).

### 5. Discussion of Findings and Recommendations for Future AD Studies

The present literature review shows that self-report measures mark the most prominent measure in HMI evaluation for AD. In comparison to that, there is further need to report accompanying behavioral measures, such as driving/TOR performance, or gaze behavior. The reason for that is quite simple: The users' reported attitudes frequently do not match the behavior that one observes [118]. For example, usability comprises effectiveness and efficiency (behavioral) as well as satisfaction (attitude) in order to get a comprehensive product evaluation. Attitudes and behavior are often separate dimensions that do not match in different various areas of research [11,119,120]. For a comprehensive evaluation of HMIs and driving automation features from a Human Factors perspective, it might be necessary to include both of these data sources. Of course, if a researcher is only interested in the user's attitudes, he/she can solely collect these, but, eventually, he/she will have to face the discussion of whether and how valid the insights are. Moreover, concerning areas of research, safety aspects received the most attention to date, followed by trust and acceptance. In the following, we discuss findings of (1) study setup, (2) data collection methods, (3) reported parameters, and (4) investigated constructs and derive recommendations (REC) for future research in the context of automated driving.

#### 5.1. Study Setup

Concerning representation and study type, we found that most studies were conducted using driving simulators. Hence, present research supports high internal validity making interpretations of effects of differing conditions on dependent measures possible. On the downside, the obtained results might lack external validity, and there is no guarantee that they generalize to real world settings, since driving simulators may lack realism due to the insufficient field of view, or motion feedback. Thus, by lacking a feeling of presence [121], the transfer of behavior found in the laboratory to real world settings might be limited [122]. Conditions in realistic driving studies are not as controlled as in a laboratory and might differ in terms of surrounding traffic, weather conditions or vehicle speeds. Despite the restriction of high internal but limited external validity, we can assume that strong and consistent effects such as the dependency of TOR response time on the driver state [13,15,21,123,124]

will also be reflected in real world driving. Especially when it comes to safety and trust issues in real driving, research should determine validity of the findings. While participants in driving simulator studies might behave in a more liberal way due to the absence of realistic severe consequences, it remains to be seen how these effects pertain to the real road. At this point, the question arises as to whether the scenarios that are tested in driving simulator research and frequency within a study are representative for human–automation interaction. Up to now, there is no valid statement on the frequency of transitions from automated to manual possible. An indication might be available from the California disengagement report [125], where developers of ADS have to report the number of safety driver interventions. As the systems are still under development and have not reached a maturity level for the commercial market, it is questionable whether these numbers resemble future series products. Future research from FOT or NDS data might be valuable for putting the importance of such scenarios into perspective. Relative validity of effects in driving simulator studies can well be assumed [126], but there are still research efforts necessary for validating safety relevant human–automation interaction scenarios. Despite this criticism, we acknowledge that driving simulators are still immersive research tools providing a certain, and, in many cases sufficient, degree of external validity [127,128]. Nevertheless, the present results point towards the need of conducting studies in real vehicles equipped with driving automation technology [50,129,130], or Wizard-of-Oz Settings [48,49,131].

⇒ REC 1: *Existing study results obtained in driving simulators should be validated in realistic on-road settings.*

Review results also revealed that most studies targeted single sessions of interaction, and thus provide only snapshots of first time use. While usability can already make reliable estimates about user behavior within a single session experiment [29], UX, trust, and acceptance might take a longer period of time until behavior and attitudes have reached a stable level [8,131,132]. Long-term studies tackling exactly these issues are rather scarce since they require high efforts. One example of such a study comes from Beggiato et al. [88] who observed users of L1 automation (i.e., Adaptive Cruise Control) over ten repeated one-hour sessions (however, due to the restriction to SAE L1, this paper would not satisfy the inclusion criteria for this review). Such studies can provide valuable insights into acceptance, trust, and system understanding. The present database only includes a small number of publications that investigated long-term use such as described in Dikmen and Burns [47]. This study example, however, did not follow a longitudinal approach, but rather a cross-sectional approach by surveying Tesla drivers. We expect important insights into behavioral adaptation over longer time periods, such as the amount of actual use or NDRTs that users engage in. Similarly to field operational tests [133] or naturalistic driving studies [134], there is still a blind spot in research on driving automation that could open up with commercial availability. First, efforts in this direction with L2 automation are reported in Gaspar and Carney [135], or planned in the L3-Pilot project [136]. Thus, there are open research directions towards ongoing and long-lasting effects of AD on use and interaction between the human operator and the ADS.

⇒ REC 2: *In contrast to single session experiments, insights into the long-lasting effects of AD usage are scarce and can benefit from longitudinal study designs.*

The age distribution of all samples within the data base showed a bell shaped curve (see Figure 4). This indicates that, across all identified studies, participants' ages were balanced and, overall, findings can be well generalized to the population. Both young and therefore novice, as well as elderly, drivers are considered in these studies. To maintain but also increase validity, we invoke authors of future studies to continue to regard users' diversity regarding age and also point to further consider gender, cultural, and other differences in their studies.

⇒ REC 3: *To increase validity, the sample characteristics should be adapted to the addressed research questions. In addition, researchers need to explain why the particular sample was chosen and considered as valid.*

## 5.2. Data Collection Methods

Concerning collection methods, the review results show that a vast majority of studies collected self-report data. In comparison to that, behavioral data were only reported in every second publication. From there, the question arises regarding what the reasons for this observation are. One obvious reason is that survey approaches [31,137] might focus more on technology readiness and deliberately collect attitudinal measures only. For these approaches, to date, there is no available behavioral criterion, such as buy/usage rates. Studies operationalizing acceptance via the TAM [35] do not provide the possibility to provide insights into behavioral measures. As soon as the functions are available, however, research needs to investigate whether predictions made by these studies hold true. With commercial availability of L2 automation, such a study could have already been conducted, but, to our knowledge, this is still missing. One positive aspect of self-report measures is that psychometrically validated scales have been applied frequently. This shows the professionalism of research of the reviewed venues and deliberate preparation of study design.

Another factor for the imbalance might be that self-report measures are much easier to collect. It does not take comparably much effort to hand out a questionnaire or interviewing participants. In contrast, the collection of behavioral data is much more complex. For example, dynamic vehicle data require extensive pre-processing before descriptive and inferential analyses can be run. The collection of eye-tracking data requires even more resources due to the need for manual calibration to ensure data quality, although such data provide the possibility to make direct inferences about cognitive processes [138]. For example, prior research has suggested that the number of gaze switches (i.e., monitoring behavior) can serve as an indicator for trust/reliance [18], or interface understanding [29]. One solution to the difficulty and extensive effort of collecting behavioral data might be experimenters' single-item ratings of interaction performance [139–141]. However, this approach requires well-trained raters and, ideally, ratings are given single-blind, so that the rater is not aware of assigned experimental conditions for the participant. In addition, there should be more than one rater to ensure inter-rater reliability. Despite requiring additional time and cost efforts, researchers should consider the collection and analysis of behavioral data should in the planning of a user study, since it can provide additional valuable insights about the tested interface or feature. It is not for nothing that the usability ISO-9241 [34] includes effectiveness and efficiency as behavioral components and satisfaction as an attitudinal component. These sources of data do not always align well [11,118], which is not necessarily a bad study outcome. It rather supports the assumption that both sources of data are necessary to derive a holistic impression of an interface. This has also been emphasized by Pettersson et al. [8,142], who expressed the urgency of triangulation in user studies. We appeal to researchers in the field of human–automation interaction to always consider additional behavioral observations. The importance of collecting behavioral measures of course depends on the respective research question. For example, there are instances where researchers are interested specifically in the public opinion of a large sample of respondents [25]. As explicitly mentioned in the Introduction of this work, we do not see the recommendations as mandatory but rather as thought-provoking. Thus, it might guide exclusive attitude research towards future work on the behavioral consequences.

⇒ REC 4: *Researchers should consider the collection of more than a single source of data (i.e., triangulating behavioral, psychophysiological, and self-reported data). Insights from self-report measures might be worthy of future research for its effects on user behavior.*

## 5.3. Parameters

The results of the literature review showed that research provides a heterogeneous pool of parameters, for example considering driving or TOR performance measures in controllability studies (see Tables 2–12). To better compare results between different laboratories, there is an urgent need to come up with a standardization of procedures when evaluating human–automation interaction. Efforts in this directions are reported in the works by Wintersberger et al. [59], which suggests TOR

performance measures based on SAE J2944 or in Naujoks et al. [60] outlining a standardized set of use cases for control transitions between levels of automation. Since there has been a lot of research on TOR controllability, it is now time to combine this body of knowledge and standardize methodology along the lines of driver distraction research [143,144]

⇒ REC 5: Consider existing or proposed standards for measurements to allow comparison of study results.

#### 5.4. Constructs

First, it turned out that some studies included parameters whose focus of research was not clearly defined. Despite setting up a large number of constructs during the expert workshop and adding further options during the review process, there still remained a considerable number of parameters that could not be assigned free of doubt. This resulted in the two constructs *general attitude* and *interaction behavior* (see Table 1). In these instances, information was rather scarce and only high level indications about interface evaluation were provided by the authors. We therefore encourage researchers to first clearly state objectives and classify these within a specific body of research. This should allow both authors and readers to eventually compare the obtained results with existing findings, and derive implications of the reported work.

⇒ REC 6: Clearly communicate which constructs are addressed by an experiment to foster transparency

The majority of parameters were well classifiable within our database. The safety construct constituted the largest part of all measures (see Table 1). This shows the importance of safety concerns when it comes to investigating driving automation technology. Investigations on TOR performance during system failures have been the first scenarios in human–automation interaction research [12–14,145]. Here, many issues such as trust [18,21,97], controllability [12], fatigue [15], or mode awareness [17] have been discovered. Right now, it seems like a re-orientation of recent research is taking place which is also reflected in the remaining constructs of the database. Trust [146] and intention to use [35] constitute precursors of actual system use, and, due to the progress in time, technology, and scientific evidence, research has more closely investigated these constructs. Here, the aforementioned open research directions of long-term studies, absolute validity concerns in realistic driving and agreeableness of attitudinal and behavioral measures apply.

Besides safety and acceptance, there remain other constructs that have rather been neglected until now. The reason why research has given less attention to usability, UX, or productivity until now might be that these are rather precursors of safety and acceptance. Another reason might be that the scenario where automation fails and humans need to step into action was an important issue to determine feasibility of driving automation in the first place. We argue that more emphasis should be paid to other types of interaction such as ongoing automation, user-initiated or planned transfers of control, as most likely, these use cases will occur more frequently than automation failures [59,147]. From that, the need to investigate efficient and effective interaction arises [29]. Furthermore, it might not be sufficient to develop interfaces that users are satisfied with [34], but which they have fun and enjoy interacting with [32,148]. Moreover, safety critical issues of AD, like TORs in SAE L3, also impact users overall driving experience [149]. Hence, an additional focus on UX and research on users' emotions and need fulfillment can pave the way to developing not only proper but great HMIs for AVs, which increase individual but also societal acceptance of this emerging technology.

We also emphasize at this point that the respective research question largely impacts the investigated constructs. The existing large number of findings on safety might benefit researchers in the formulation of research questions since statements on future research and unanswered questions are an inherent part of a researcher's work. At the same time, this implies that there is a lot of room for research questions beyond established construct.

⇒ REC 7: Depending on the particular research question, we suggest going beyond established constructs like safety, trust, acceptance, or workload and regard topics from different perspectives.



## 6. Limitations

The literature review presented here comes with some limitations. First, the restriction on the most relevant six sources (three journals and and three conferences) limits the reach of included works, and also other venues/journals publish cutting-edge research on driving automation matching our inclusion criteria. However, comparing the obtained results with a prior review concentrating only on one of the six included sources [150] shows that the results did not drastically change after extension. Second, we (subjectively) felt a trend to experiments that are published multiple times with slightly adapted focus (for example, a conference publication presenting first insights into an experiment is followed by a more detailed journal submission). As our review focused on publications rather than single experiments, we cannot guarantee that some studies in our database are duplicates, which may have slightly impacted the results (for example regarding the distribution of participants' age, study types, or the share of the levels of automation addressed). Third, the inter-rater reliability is calculated based on the inclusion/exclusion criteria of papers, while the subsequent classification process is not completely free from subjective assessment. Although we tried to keep the level of subjective interpretation to a minimum (by defining a standardized reviewing procedure), not all involved decisions were free from ambiguity. For example, some publications evaluated multiple constructs but did not provide information on the mapping of investigated parameters to these constructs. In such cases, the decision is burdened on the reviewer, and, to minimize such effects, the authors discussed potential inconsistencies together and made adaptations (if necessary) before the final database analyses. Moreover, the methodological insights presented here combine a large variety of research questions spanning over many different areas of research. To derive specific insights for one target of research (e.g., how do drivers take over from automated to manual control), a separate consideration is necessary using relevant publications for such a category. Despite this limitation, our results mark a first step into this direction by providing researchers with methods and parameters that they most likely have to consider when setting up an experiment investigating a certain construct with relevance to automated driving. Eventually, this work contains literature from six venues and one needs to be aware that the recommendations apply to the literature that was included here. However, the claim that these venues cover a representative share of driving automation studies strengthens the argument that they might apply for venues beyond. The initially applied keywords to reduce the amount of papers might be criticized as arbitrary and a different set of keywords could have led to a different collection of publications. However, these keywords were chosen after extensive discussion among the authors of this paper and thus are based on the best of our knowledge of human factors research in driving automation.

## 7. Conclusions

In this article, we have reviewed the status quo of methods utilized in human factors research in driving automation. We followed a structured approach to give an overview of the research domain by selecting relevant papers, and reviewing them in a standardized manner using a relational database. There is a good portion of research in different aspects of driving automation indicating that researchers in the community work on the issue of developing and improving human-machine interfaces for automated vehicles. When researchers plan to engage in research and development of automated driving, the present work provides them with an overview of the current landscape. Thus, one can derive information about main research areas and emerging trends that have not been studied extensively yet. Additionally, this literature review provides researchers and practitioners with suggestions about methodological tools (i.e., collection methods and specific parameters) that they can use when assessing a certain construct of a driving automation system. To conclude, we list a set of recommendations to be considered in future experiments addressing automated driving:



- **Existing study results obtained in driving simulators should be validated in realistic on-road settings.** As most experimental results were obtained in driving simulators (or used even lower degree of realism/immersion), their main findings must urgently be validated in more realistic settings, especially when addressing constructs that incorporate risk (such as trust in automation).
- **In contrast to single session experiments, insights into the long-lasting effects of AD usage are scarce and can benefit from longitudinal study designs.** Another huge potential for future work are longitudinal studies. Such studies cannot only validate results obtained in single-session experiments: they might even reveal new issues which have not been addressed yet.
- **Depending on the particular research question, we suggest going beyond established constructs like safety, trust, acceptance, or workload.** In addition, take constructs into account that have not yet been intensively addressed (such as personalization, cooperation, wellbeing, etc.). Cooperation in particular refers to different parties performing as a team together rather than only one party at a time [87]. When designing HMIs for vehicles, consider the full spectrum of user experience research and include user satisfaction (such as hedonic qualities) in HMI evaluation also as these aspects will finally decide between success or defeat of the technology on the market.
- **To increase validity, the sample characteristics should be adapted to the addressed research questions. In addition, researchers need to explain why the particular sample was chosen and considered valid.** Aim for a participant sample that allows for investigating the proposed research question plausibly. Do not only list a more diverse sample as a limitation, but also discuss how your results could be affected by biases in participant sampling.
- **Researchers should consider the collection of more than a single source of data (i.e., triangulating behavioral, psychophysiological, and self-reported data). Insights from self-report measures might be worthy of future research for its effects on user behavior.** Evaluation of different types of data sometimes leads to contradicting results. Such conflicts should not be avoided, as in the best case they contribute to a better understanding of established theory. Only comprehensive evaluation of the involved factors allows for drawing meaningful conclusions.
- **Consider existing or proposed standards for measurements to allow for comparison of study results.** The possibility to compare study results is a key element of scientific practices. Thus, if possible, utilize standardized methods (regarding parameters, their measurement times, as well as their calculation). In case there is no such standard or best practice, instead of inventing additional measurements, build upon related work.
- **Clearly communicate which constructs are addressed by an experiment to foster transparency.** Minimize the potential for ambiguity. Clearly state which constructs are investigated, and how the selection of methods/parameters used for evaluation are related to them.

**Author Contributions:** Conceptualization, A.-K.F., Y.F., and P.W.; methodology, A.-K.F., Y.F., P.W., and V.G.; formal analysis, A.-K.F. and P.W.; writing—original draft preparation, A.-K.F., Y.F., and P.W.; writing—review and editing, A.-K.F., Y.F., P.W., and A.R.; visualization, A.-K.F.; supervision, A.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. SAE. *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems*; SAE International: Warrendale, PE, USA, 2018.
2. Sarter, N.B.; Woods, D.D.; Billings, C.E. Automation surprises. *Handb. Hum. Factors Ergon.* **1997**, *2*, 1926–1943.
3. Billings, C.E. *Aviation Automation: The Search for a Human-Centered Approach*; CRC Press: Boca Raton, FL, USA, 2018.
4. Pritchett, A.R. Aviation automation: General perspectives and specific guidance for the design of modes and alerts. *Rev. Hum. Factors Ergon.* **2009**, *5*, 82–113. [[CrossRef](#)]

5. AdaptiVe Consortium. *Final Human Factors Recommendations (D3.3)*; 2017. Available online: <http://eprints.whiterose.ac.uk/161983/> (accessed on 15 November 2020).
6. RESPONSE Consortium. *Code of Practice for the Design and Evaluation of ADAS: RESPONSE 3: A PREVENT Project*; 2006. Available online: [https://www.acea.be/uploads/publications/20090831\\_Code\\_of\\_Practice\\_ADAS.pdf](https://www.acea.be/uploads/publications/20090831_Code_of_Practice_ADAS.pdf) (accessed on 15 November 2020).
7. Hoeger, R.; Zeng, H.; Hoess, A.; Kranz, T.; Boverie, S.; Strauss, M. *The Future of Driving—HAVEit (Final Report, Deliverable D61.1)*; Continental Automotive GmbH: Regensburg, Germany, 2011.
8. Pettersson, I.; Lachner, F.; Frison, A.K.; Riener, A.; Butz, A. A Bermuda Triangle? A Review of Method Application and Triangulation in User Experience Evaluation. In Proceedings of the 2018 CHI Conference, Montreal, QC, Canada, 21–26 April 2018; pp. 1–16. [[CrossRef](#)]
9. Lamm, L.; Wolff, C. Exploratory Analysis of the Research Literature on Evaluation of In-Vehicle Systems. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Utrecht, The Netherlands, 22–25 September 2019; ACM: New York, NY, USA, 2019; pp. 60–69.
10. Ayoub, J.; Zhou, F.; Bao, S.; Yang, X.J. From Manual Driving to Automated Driving: A Review of 10 Years of AutoUI. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Utrecht, The Netherlands, 22–25 September 2019; ACM: New York, NY, USA, 2019; pp. 70–90.
11. Forster, Y.; Hergeth, S.; Naujoks, F.; Krems, J.F. How Usability can Save the Day: Methodological Considerations for Making Automated Driving a Success Story. In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Toronto, ON, Canada, 23–25 September 2018.
12. Naujoks, F.; Mai, C.; Neukum, A. The effect of urgency of take-over requests during highly automated driving under distraction conditions. *Adv. Hum. Asp. Transp.* **2014**, *7*, 431.
13. Gold, C.; Damböck, D.; Lorenz, L.; Bengler, K. “Take over!” How long does it take to get the driver back into the loop? In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, San Diego, CA, USA, 30 September–4 October 2013; SAGE Publications: Los Angeles, CA, USA, 2013.
14. Merat, N.; Jamson, A.H.; Lai, F.C.; Daly, M.; Carsten, O.M. Transition to manual: Driver behavior when resuming control from a highly automated vehicle. *Transp. Res. Part Traffic Psychol. Behav.* **2014**, *27*, 274–282. [[CrossRef](#)]
15. Jarosch, O.; Kuhnt, M.; Paradies, S.; Bengler, K. It’s Out of Our Hands Now! Effects of Non-Driving Related Tasks During Highly Automated Driving on Drivers’ Fatigue. In Proceedings of the 9th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Manchester Village, VT, USA, 26–29 June 2017.
16. Neubauer, C.; Matthews, G.; Langheim, L.; Saxby, D. Fatigue and voluntary utilization of automation in simulated driving. *Hum. Factors* **2012**, *54*, 734–746. [[CrossRef](#)]
17. Feldhütter, A.; Segler, C.; Bengler, K. Does Shifting Between Conditionally and Partially Automated Driving Lead to a Loss of Mode Awareness? In Proceedings of the International Conference on Applied Human Factors and Ergonomics, Los Angeles, CA, USA, 17–21 July 2017.
18. Hergeth, S.; Lorenz, L.; Vilimek, R.; Krems, J.F. Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving. *Hum. Factors* **2016**, *58*, 509–519. [[CrossRef](#)]
19. Hergeth, S.; Lorenz, L.; Krems, J.F. Prior familiarization with takeover requests affects drivers’ takeover performance and automation trust. *Hum. Factors* **2017**, *59*, 457–470. [[CrossRef](#)]
20. Forster, Y.; Naujoks, F.; Neukum, A. Increasing anthropomorphism and trust in automated driving functions by adding speech output. In Proceedings of the Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017.
21. Wintersberger, P.; Riener, A.; Schartmüller, C.; Frison, A.K.; Weigl, K. Let Me Finish before I Take Over: Towards Attention Aware Device Integration in Highly Automated Vehicles. In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Toronto, ON, Canada, 23–25 September 2018; ACM: New York, NY, USA, 2018; pp. 53–65.
22. Merat, N.; Lee, J.D. Preface to the special section on human factors and automation in vehicles: Designing highly automated vehicles with the driver in mind. *Hum. Factors* **2012**, *54*, 681–686. [[CrossRef](#)]

23. Nordhoff, S.; de Winter, J.; Madigan, R.; Merat, N.; van Arem, B.; Happee, R. User acceptance of automated shuttles in Berlin-Schöneberg: A questionnaire study. *Transp. Res. Part Traffic Psychol. Behav.* **2018**, *58*, 843–854. [[CrossRef](#)]
24. Osswald, S.; Wurhofer, D.; Trösterer, S.; Beck, E.; Tscheligi, M. Predicting information technology usage in the car: Towards a car technology acceptance model. In Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Portsmouth, NH, USA, 17–19 October 2012.
25. Kyriakidis, M.; Happee, R.; de Winter, J.C.F. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transp. Res. Part Traffic Psychol. Behav.* **2015**, *32*, 127–140. [[CrossRef](#)]
26. Wintersberger, P.; Riener, A.; Frison, A.K. Automated Driving System, Male, or Female Driver: Who'd You Prefer? Comparative Analysis of Passengers' Mental Conditions, Emotional States & Qualitative Feedback. In Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ann Arbor, MI, USA, 24–26 October 2016; ACM: New York, NY, USA, 2016; pp. 51–58.
27. Payre, W.; Cestac, J.; Dang, N.T.; Vienne, F.; Delhomme, P. Impact of training and in-vehicle task performance on manual control recovery in an automated car. *Transp. Res. Part Traffic Psychol. Behav.* **2017**, *46*, 216–227. [[CrossRef](#)]
28. Wintersberger, P.; Frison, A.K.; Riener, A.; von Sawitzky, T. Fostering User Acceptance and Trust in Fully Automated Vehicles: Evaluating the Potential of Augmented Reality. *Presence* **2019**, *27*, 1–17. [[CrossRef](#)]
29. Forster, Y.; Hergeth, S.; Naujoks, F.; Beggiato, M.; Krems, J.F.; Keinath, A. Learning to Use Automation: Behavioral Changes in Interaction with Automated Driving Systems. *Transp. Res. Part Traffic Psychol. Behav.* **2019**, *62*, 599–614. [[CrossRef](#)]
30. Naujoks, F.; Hergeth, S.; Wiedemann, K.; Schömig, N.; Forster, Y.; Keinath, A. Test procedure for evaluating the human-machine interface of vehicles with automated driving systems. *Traffic Inj. Prev.* **2019**, *20*, S146–S151. [[CrossRef](#)]
31. Rödel, C.; Stadler, S.; Meschtscherjakov, A.; Tscheligi, M. Towards Autonomous Cars. In Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Seattle, WA, USA, 17–19 September 2014; pp. 1–8. [[CrossRef](#)]
32. Frison, A.K.; Wintersberger, P.; Riener, A.; Schartmüller, C. Driving Hotzenplotz: A Hybrid Interface for Vehicle Control Aiming to Maximize Pleasure in Highway Driving. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany, 24–27 September 2017; pp. 236–244.
33. Frison, A.K.; Wintersberger, P.; Riener, A.; Schartmüller, C.; Boyle, L.; Miller, E.; Weigl, K. In UX We Trust: Investigation of Aesthetics and Usability of Driver-Vehicle Interfaces and Their Impact on the Perception of Automated Driving. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; pp. 1–13.
34. DIN-EN ISO. *Ergonomie der Mensch-System-Interaktion—Teil 210: Prozess zur Gestaltung Gebrauchstauglicher Interaktiver Systeme*; EIN e.V.: Berlin, Germany, 2011.
35. Davis, F.D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **1989**, *13*, 319. [[CrossRef](#)]
36. Venkatesh, V.; Morris, M.G.; Davis, G.B.; Davis, F.D. User acceptance of information technology: Toward a unified view. *MIS Q.* **2003**, *27*, 425–478. [[CrossRef](#)]
37. Hassenzahl, M.; Beu, A.; Burmester, M. Engineering joy. *IEEE Softw.* **2001**, *18*, 70–76. [[CrossRef](#)]
38. Sheldon, K.M.; Elliot, A.J.; Kim, Y.; Kasser, T. What is satisfying about satisfying events? Testing 10 candidate psychological needs. *J. Personal. Soc. Psychol.* **2001**, *80*, 325. [[CrossRef](#)]
39. Hassenzahl, M.; Diefenbach, S.; Göritz, A. Needs, affect, and interactive products—Facets of user experience. *Interact. Comput.* **2010**, *22*, 353–362. [[CrossRef](#)]
40. Frison, A.K.; Wintersberger, P.; Liu, T.; Riener, A. Why do you like to drive automated?: A context-dependent analysis of highly automated driving to elaborate requirements for intelligent user interfaces. In Proceedings of the 24th International Conference on Intelligent User Interfaces, Los Angeles, CA, USA, 17–20 March 2019; ACM: New York, NY, USA, 2019; pp. 528–537.

41. Forster, Y.; Naujoks, F.; Neukum, A. Your Turn or My Turn? Design of a Human–Machine Interface for Conditional Automation. In Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ann Arbor, MI, USA, 24–26 October 2016; pp. 253–260. [[CrossRef](#)]
42. Louw, T.; Madigan, R.; Carsten, O.; Merat, N. Were they in the loop during automated driving? Links between visual attention and crash potential. *Inj. Prev.* **2017**, *23*, 281–286. [[CrossRef](#)]
43. Donmez, B.; Boyle, L.N.; Lee, J.D.; McGehee, D.V. Drivers' attitudes toward imperfect distraction mitigation strategies. *Transp. Res. Part Traffic Psychol. Behav.* **2006**, *9*, 387–398. [[CrossRef](#)]
44. van Veen, T.; Karjanto, J.; Terken, J. Situation awareness in automated vehicles through proximal peripheral light signals. In Proceedings of the 8th International Conference on Automotive User Interfaces and Vehicular Applications, Ann Arbor, MI, USA, 24–26 October 2016; ACM: New York, NY, USA, 2016; pp. 287–292.
45. Yusof, N.M.; Karjanto, J.; Terken, J.; Delbressine, F.; Hassan, M.Z.; Rauterberg, M. The Exploration of Autonomous Vehicle Driving Styles. In Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ann Arbor, MI, USA, 24–26 October 2016; ACM: New York, NY, USA, 2016; pp. 245–252. [[CrossRef](#)]
46. Trösterer, S.; Meschtscherjakov, A.; Mirnig, A.G.; Lupp, A.; Gärtner, M.; McGee, F.; McCall, R.; Tscheligi, M.; Engel, T. What We Can Learn from Pilots for Handovers and (De)Skilling in Semi-Autonomous Driving. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany, 24–27 September 2017; ACM: New York, NY, USA, 2017; pp. 173–182. [[CrossRef](#)]
47. Dikmen, M.; Burns, C.M. Autonomous driving in the real world: Experiences with tesla autopilot and summon. In Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ann Arbor, MI, USA, 24–26 October 2016; ACM: New York, NY, USA, 2016; pp. 225–228.
48. Omozik, K.; Yang, Y.; Kuntermann, I.; Hergeth, S.; Bengler, K. How long does it take to relax? Observation of driver behaviors during real-world conditionally automated driving. In Proceedings of the 10th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Santa Fe, NM, USA, 24–27 June 2019.
49. Naujoks, F.; Purucker, C.; Wiedemann, K.; Marberger, C. Noncritical State Transitions During Conditionally Automated Driving on German Freeways: Effects of Non–Driving Related Tasks on Takeover Time and Takeover Quality. *Hum. Factors* **2019**, *61*, 1–18. [[CrossRef](#)] [[PubMed](#)]
50. Wintersberger, P.; Frison, A.K.; Riener, A. Man vs. Machine: Comparing a Fully Automated Bus Shuttle with a Manually Driven Group Taxi in a Field Study. In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Toronto, ON, Canada, 23–25 September 2018; ACM: New York, NY, USA, 2018; pp. 215–220.
51. Reig, S.; Norman, S.; Morales, C.G.; Das, S.; Steinfeld, A.; Forlizzi, J. A Field Study of Pedestrians and Autonomous Vehicles. In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI'18), Toronto, ON, Canada, 23–25 September 2018; ACM: New York, NY, USA, 2018; pp. 198–209. [[CrossRef](#)]
52. Kunze, A.; Summerskill, S.; Marshall, R.; Filtner, A. Augmented Reality Displays for Communicating Uncertainty Information in Automated Driving. In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Toronto, ON, Canada, 23–25 September 2018; ACM Press: New York, NY, USA, 2018; pp. 164–175.
53. Hock, P.; Kraus, J.; Walch, M.; Lang, N.; Baumann, M. 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, Ann Arbor, MI, USA, 24–26 October 2016; pp. 105–112. [[CrossRef](#)]
54. Chang, C.M.; Toda, K.; Sakamoto, D.; Igarashi, T. Eyes on a Car: An Interface Design for Communication between an Autonomous Car and a Pedestrian. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany, 24–27 September 2017; ACM: New York, NY, USA, 2017; pp. 65–73. [[CrossRef](#)]

55. van der Meulen, H.; Kun, A.L.; Janssen, C.P. Switching Back to Manual Driving: How Does it Compare to Simply Driving Away After Parking? In Proceedings of the 8th International Conference on Automotive User Interfaces and Vehicular Applications, Ann Arbor, MI, USA, 24–26 October 2016; ACM: New York, NY, USA, 2016; pp. 229–236.
56. Sikkenk, M.; Terken, J. Rules of conduct for autonomous vehicles. In Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Nottingham, UK, 1–3 September 2015; ACM: New York, NY, USA, 2015; pp. 19–22.
57. Balters, S.; Sibi, S.; Johns, M.; Steinert, M.; Ju, W. Learning-by-Doing: Using Near Infrared Spectroscopy to Detect Habituation and Adaptation in Automated Driving. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany, 24–27 September 2017; ACM: New York, NY, USA, 2017; pp. 134–143. [[CrossRef](#)]
58. McCall, R.; McGee, F.; Meschtscherjakov, A.; Louveton, N.; Engel, T. Towards a taxonomy of autonomous vehicle handover situations. In Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ann Arbor, MI, USA, 24–26 October 2016; ACM: New York, NY, USA, 2016; pp. 193–200.
59. Wintersberger, P.; Green, P.; Riemer, A. Am I driving or are you or are we both? A taxonomy for handover and handback in automated driving. In Proceedings of the 9th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Manchester Village, VT, USA, 26–29 June 2017.
60. Naujoks, F.; Hergeth, S.; Wiedemann, K.; Schömig, N.; Keinath, A. Use Cases for Assessing, Testing, and Validating the Human Machine Interface of Automated Driving Systems. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Philadelphia, PA, USA, 1–5 October 2018; Sage Publications: Los Angeles, CA, USA, 2018; pp. 1873–1877.
61. Gold, C.; Naujoks, F.; Radlmayr, J.; Bellem, H.; Jarosch, O. Testing Scenarios for Human Factors Research in Level 3 Automated Vehicles. In Proceedings of the International Conference on Applied Human Factors and Ergonomics, Los Angeles, CA, USA, 17–21 July 2017; Springer: Berlin/Heidelberg, Germany; Los Angeles, CA, USA, 2017.
62. Forster, Y.; Hergeth, S.; Naujoks, F.; Krems, J.F. Unskilled and Unaware: Subpar Users of Automated Driving Systems Make Spurious Decisions. In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Toronto, ON, Canada, 23–25 September 2018; ACM: New York, NY, USA, 2018; pp. 159–163.
63. Millonig, A.; Fröhlich, P. Where Autonomous Buses Might and Might Not Bridge the Gaps in the 4 A's of Public Transport Passenger Needs: a Review. In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Toronto, ON, Canada, 23–25 September 2018; ACM: New York, NY, USA, 2018; pp. 291–297.
64. Inners, M.; Kun, A.L. Beyond liability: Legal issues of human-machine interaction for automated vehicles. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany, 24–27 September 2017; ACM: New York, NY, USA, 2017; pp. 245–253.
65. Roider, F.; Rümelin, S.; Pflöging, B.; Gross, T. The effects of situational demands on gaze, speech and gesture input in the vehicle. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany, 24–27 September 2017; ACM: New York, NY, USA, 2017; pp. 94–102.
66. Merenda, C.; Kim, H.; Gabbard, J.L.; Leong, S.; Large, D.R.; Burnett, G. Did You See Me?: Assessing Perceptual vs. Real Driving Gains Across Multi-Modal Pedestrian Alert Systems. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany, 24–27 September 2017; ACM: New York, NY, USA, 2017; pp. 40–49.
67. Liu, R.; Kwak, D.; Devarakonda, S.; Bekris, K.; Iftode, L. Investigating remote driving over the LTE network. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany, 24–27 September 2017; ACM: New York, NY, USA, 2017; pp. 264–269.
68. Knappe, G.; Keinath, A.; Meinecke, C. Empfehlungen für die Bestimmung der Spurhaltegröße im Kontext der Fahrsimulation. *MMI-Interakt.* **2006**, *11*, 3–13.
69. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*; Elsevier: Amsterdam, The Netherlands, 1988; Volume 52, pp. 139–183.



70. Lee, J.; Kim, N.; Imm, C.; Kim, B.; Yi, K.; Kim, J. A question of trust: An ethnographic study of automated cars on real roads. In Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ann Arbor, MI, USA, 24–26 October 2016; ACM: New York, NY, USA, 2016; pp. 201–208.
71. Currano, R.; Park, S.Y.; Domingo, L.; Garcia-Mancilla, J.; Santana-Mancilla, P.C.; Gonzalez, V.M.; Ju, W.; Vámos, I. Observations of Pedestrian Interactions with Driverless Cars in Mexico. In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Toronto, ON, Canada, 23–25 September 2018; ACM: New York, NY, USA, 2018; pp. 210–220.
72. Telpaz, A.; Rhindress, B.; Zelman, I.; Tsimhoni, O. Haptic seat for automated driving: preparing the driver to take control effectively. In Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Nottingham, UK, 1–3 September 2015; ACM: New York, NY, USA, 2015; pp. 23–30.
73. Frison, A.K.; Aigner, L.; Wintersberger, P.; Riener, A. Who is Generation A?: Investigating the Experience of Automated Driving for Different Age Groups. In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Toronto, ON, Canada, 23–25 September 2018; ACM: New York, NY, USA, 2018; pp. 94–104.
74. Forster, Y.; Kraus, J.; Feinauer, S.; Baumann, M. Calibration of Trust Expectancies in Conditionally Automated Driving by Brand, Reliability Information and Introductory Videos: An Online Study. In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI'18), Toronto, ON, Canada, 23–25 September 2018; ACM: New York, NY, USA, 2018; pp. 118–128. [[CrossRef](#)]
75. Bashiri, B.; Mann, D.D. Drivers' mental workload in agricultural semi-autonomous vehicles. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, San Diego, CA, USA, 30 September–4 October 2013; SAGE Publications: Los Angeles, CA, USA, 2013; Volume 57, pp. 1795–1799.
76. Biondi, F.N.; Lohani, M.; Hopman, R.; Mills, S.; Cooper, J.M.; Strayer, D.L. 80 MPH and out-of-the-loop: Effects of real-world semi-automated driving on driver workload and arousal. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Philadelphia, PA, USA, 1–5 October 2018; SAGE Publications: Los Angeles, CA, USA, 2018; Volume 62, pp. 1878–1882.
77. Maurer, S.; Erbach, R.; Kraiem, I.; Kuhnert, S.; Grimm, P.; Rukzio, E. Designing a Guardian Angel: Giving an Automated Vehicle the Possibility to Override Its Driver. In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI'18), Toronto, ON, Canada, 23–25 September 2018.
78. Clark, H.; McLaughlin, A.C.; Feng, J. Situational Awareness and Time to Takeover: Exploring an Alternative Method to Measure Engagement with High-Level Automation. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Austin, TX, USA, 9–13 October 2017; SAGE Publications: Los Angeles, CA, USA, 2017; Volume 61, pp. 1452–1456.
79. Karjanto, J.; Yusof, N.M.; Wang, C.; Terken, J.; Delbressine, F.; Rauterberg, M. The effect of peripheral visual feedforward system in enhancing situation awareness and mitigating motion sickness in fully automated driving. *Transp. Res. Part Traffic Psychol. Behav.* **2018**, *58*, 678–692. [[CrossRef](#)]
80. Naujoks, F.; Höfling, S.; Purucker, C.; Zeeb, K. From partial and high automation to manual driving: Relationship between non-driving related tasks, drowsiness and take-over performance. *Accid. Anal. Prev.* **2018**, *121*, 28–42. [[CrossRef](#)]
81. Oliveira, L.; Luton, J.; Iyer, S.; Burns, C.; Mouzakitis, A.; Jennings, P.; Birrell, S. Evaluating How Interfaces Influence the User Interaction with Fully Autonomous Vehicles. In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI'18), Toronto, ON, Canada, 23–25 September 2018; ACM: New York, NY, USA, 2018; pp. 320–331. [[CrossRef](#)]
82. Reimer, B.; Pettinato, A.; Fridman, L.; Lee, J.; Mehler, B.; Seppelt, B.; Park, J.; Iagnemma, K. Behavioral Impact of Drivers' Roles in Automated Driving. In Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (Automotive'UI 16), Ann Arbor, MI, USA, 24–26 October 2016; ACM: New York, NY, USA, 2016; pp. 217–224. [[CrossRef](#)]
83. Bellem, H.; Klüver, M.; Schrauf, M.; Schöner, H.P.; Hecht, H.; Krems, J.F. Can we study autonomous driving comfort in moving-base driving simulators? A validation study. *Hum. Factors* **2017**, *59*, 442–456. [[CrossRef](#)]



84. Terken, Z.; Haex, R.; Beurgens, L.; Arslanova, E.; Vrachni, M.; Terken, J.; Szostak, D. Unwinding After Work: An In-car Mood Induction System for Semi-autonomous Driving. In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI'13), Eindhoven, The Netherlands, 28–30 October 2013; ACM: New York, NY, USA, 2013; pp. 246–249. [[CrossRef](#)]
85. Glatz, C.; Krupenia, S.S.; Bülthoff, H.H.; Chuang, L.L. Use the right sound for the right job: Verbal commands and auditory icons for a task-management system favor different information processes in the brain. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; ACM: New York, NY, USA, 2018; p. 472.
86. Guo, C.; Sentouh, C.; Popieul, J.C.; Haué, J.B.; Langlois, S.; Loeillet, J.J.; Soualmi, B.; That, T.N. Cooperation between driver and automated driving system: Implementation and evaluation. *Transp. Res. Part Traffic Psychol. Behav.* **2017**. [[CrossRef](#)]
87. Walch, M.; Sieber, T.; Hock, P.; Baumann, M.; Weber, M. 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, Ann Arbor, MI, USA, 24–26 October 2016; ACM: New York, NY, USA, 2016; pp. 261–268.
88. Beggiato, M.; Pereira, M.; Petzoldt, T.; Krems, J.F. Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study. *Transp. Res. Part Traffic Psychol. Behav.* **2015**, *35*, 75–84. [[CrossRef](#)]
89. Mattes, S. The lane-change-task as a tool for driver distraction evaluation. *Qual. Work. Prod. Enterp. Future* **2003**, *57*, 60.
90. Naujoks, F.; Purucker, C.; Neukum, A.; Wolter, S.; Steiger, R. Controllability of partially automated driving functions—does it matter whether drivers are allowed to take their hands off the steering wheel? *Transp. Res. Part Traffic Psychol. Behav.* **2015**, *35*, 185–198. [[CrossRef](#)]
91. Jian, J.Y.; Bisantz, A.M.; Drury, C.G. Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cogn. Ergon.* **2000**, *4*, 53–71. [[CrossRef](#)]
92. Rotter, J.B. A new scale for the measurement of interpersonal trust 1. *J. Personal.* **1967**, *35*, 651–665. [[CrossRef](#)]
93. McKnight, D.H.; Choudhury, V.; Kacmar, C. Developing and validating trust measures for e-commerce: An integrative typology. *Inf. Syst. Res.* **2002**, *13*, 334–359. [[CrossRef](#)]
94. Van Der Laan, J.D.; Heino, A.; De Waard, D. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transp. Res. Part Emerg. Technol.* **1997**, *5*, 1–10. [[CrossRef](#)]
95. Schaefer, K.E. Measuring trust in human robot interactions: Development of the “trust perception scale-HRI”. In *Robust Intelligence and Trust in Autonomous Systems*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 191–218.
96. Merritt, S.M.; Heimbaugh, H.; LaChapell, J.; Lee, D. I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Hum. Factors* **2013**, *55*, 520–534. [[CrossRef](#)] [[PubMed](#)]
97. Helldin, T.; Falkman, G.; Riveiro, M.; Davidsson, S. 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, Eindhoven, The Netherlands, 28–30 October 2013; ACM: New York, NY, USA, 2013; pp. 210–217.
98. Pauzié, A. A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intell. Transp. Syst.* **2008**, *2*, 315–322. [[CrossRef](#)]
99. Zijlstra, F.R.H. Efficiency in Work Behaviour: A Design Approach for Modern Tools. Ph.D. Dissertation, Delft Technical University, Delft University Press, Delft, The Netherlands 1993.
100. Eilers, K.; Nachreiner, F.; Hänecke, K. Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung. *Z. Arbeitswissenschaft* **1986**, *4*, 214–224.
101. Wierwille, W.W.; Casali, J.G. A validated rating scale for global mental workload measurement applications. In Proceedings of the Human Factors society Annual Meeting, Norfolk, VI, USA, 10–14 October 1983; Sage Publications: Los Angeles, CA, USA, 1983; Volume 27, pp. 129–133.
102. Endsley, M.R. Situation awareness global assessment technique (SAGAT). In Proceedings of the IEEE 1988 National Aerospace and Electronics Conference, Dayton, OH, USA, 23–27 May 1988; pp. 789–795.

103. Helton, W.S. Validation of a short stress state questionnaire. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, New Orleans, LA, USA, 20–24 September 2004; SAGE Publications: Los Angeles, CA, USA, 2004; Volume 48, pp. 1238–1242.
104. Matthews, G.; Szalma, J.; Panganiban, A.R.; Neubauer, C.; Warm, J.S. Profiling task stress with the dundee stress state questionnaire. *Psychol. Stress. New Res.* **2013**, *1*, 49–90.
105. Matthews, G.; Desmond, P.A.; Joyner, L.; Carcary, B.; Gilliland, K. Validation of the driver stress inventory and driver coping questionnaire. In Proceedings of the International Conference on Traffic and Transport Psychology, Valencia, Spain, 22–25 May 1996; pp. 1–27.
106. Deb, S.; Strawderman, L.; DuBien, J.; Smith, B.; Carruth, D.W.; Garrison, T.M. Evaluating pedestrian behavior at crosswalks: Validation of a pedestrian behavior questionnaire for the US population. *Accid. Anal. Prev.* **2017**, *106*, 191–201. [[CrossRef](#)]
107. Hoyle, R.H.; Stephenson, M.T.; Palmgreen, P.; Lorch, E.P.; Donohew, R.L. Reliability and validity of a brief measure of sensation seeking. *Personal. Individ. Differ.* **2002**, *32*, 401–414. [[CrossRef](#)]
108. Ajzen, I. The theory of planned behavior. *Organ. Behav. Hum. Decis. Process.* **1991**, *50*, 179–211. [[CrossRef](#)]
109. Shahid, A.; Wilkinson, K.; Marcu, S.; Shapiro, C.M. Karolinska sleepiness scale (KSS). In *STOP, THAT and One Hundred Other Sleep Scales*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 209–210.
110. Matthews, G.; Desmond, P.A.; Joyner, L.; Carcary, B.; Gilliland, K. A comprehensive questionnaire measure of driver stress and affect. *Traffic Transp. Psychol. Theory Appl.* **1997**, 317–324.
111. Smets, E.; Garssen, B.; Bonke, B.D.; De Haes, J. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *J. Psychosom. Res.* **1995**, *39*, 315–325. [[CrossRef](#)]
112. Kujala, S.; Roto, V.; Väänänen-Vainio-Mattila, K.; Karapanos, E.; Sinnelä, A. UX Curve: A method for evaluating long-term user experience. *Interact. Comput.* **2011**, *23*, 473–483. [[CrossRef](#)]
113. Hassenzahl, M.; Burmester, M.; Koller, F. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & Computer 2003*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 187–196.
114. Laugwitz, B.; Held, T.; Schrepp, M. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and Usability Engineering Group*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 63–76.
115. Rice, S.; Winter, S. A quick affect scale: Providing evidence for validity and reliability. In Proceedings of the 10th International Conference on Interdisciplinary Social Sciences, Split, Croatia, 11–14 June 2015.
116. Kennedy, R.S.; Lane, N.E.; Berbaum, K.S.; Lilienthal, M.G. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *Int. J. Aviat. Psychol.* **1993**, *3*, 203–220. [[CrossRef](#)]
117. Gianaros, P.J.; Muth, E.R.; Mordkoff, J.T.; Levine, M.E.; Stern, R.M. A questionnaire for the assessment of the multiple dimensions of motion sickness. *Aviat. Space Environ. Med.* **2001**, *72*, 115.
118. Nielsen, J.; Levy, J. Measuring usability: Preference vs. performance. *Commun. ACM* **1994**, *37*, 66–76. [[CrossRef](#)]
119. Endsley, M.R. The divergence of objective and subjective situation awareness: A meta-analysis. *J. Cogn. Eng. Decis. Mak.* **2020**, *14*, 34–53. [[CrossRef](#)]
120. Hancock, P.A.; Matthews, G. Workload and performance: Associations, insensitivities, and dissociations. *Hum. Factors* **2019**, *61*, 374–392. [[CrossRef](#)]
121. Slater, M. A note on presence terminology. *Presence Connect* **2003**, *3*, 1–5.
122. Will, S. Development of a Presence Model for Driving Simulators Based on Speed Perception in a Motorcycle Riding Simulator. Ph.D. Thesis, University of Wuerzburg, Wuerzburg, Germany, 2017.
123. Naujoks, F.; Befelein, D.; Wiedemann, K.; Neukum, A. A review of non-driving-related tasks used in studies on automated driving. In Proceedings of the International Conference on Applied Human Factors and Ergonomics, Los Angeles, CA, USA, 17–21 July 2017; Springer: Los Angeles, CA, USA, 2017; pp. 525–537.
124. Wandtner, B. Non-Driving Related Tasks in Highly Automated Driving—Effects of Task Characteristics and Drivers’ Self-Regulation on Take-Over Performance. Ph.D. Thesis, University of Wuerzburg, Wuerzburg, Germany, 2018.
125. State of California. *Disengagement Report*; Department of Motor Vehicles: Sacramento, CA, USA, 2017.
126. Fisher, D.L.; Rizzo, M.; Caird, J.; Lee, J.D. *Handbook of Driving Simulation for Engineering, Medicine, and Psychology*; CRC Press: Boca Raton, FL, USA, 2011.

127. Kemeny, A.; Panerai, F. Evaluating perception in driving simulation experiments. *Trends Cogn. Sci.* **2003**, *7*, 31–37. [[CrossRef](#)]
128. Hock, P.; Kraus, J.; Babel, F.; Walch, M.; Rukzio, E.; Baumann, M. How to Design Valid Simulator Studies for Investigating User Experience in Automated Driving: Review and Hands-On Considerations. In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Toronto, ON, Canada, 23–25 September 2018; ACM: New York, NY, USA, 2018; pp. 105–117.
129. Banks, V.A.; Eriksson, A.; O'Donoghue, J.; Stanton, N.A. Is partially automated driving a bad idea? Observations from an on-road study. *Appl. Ergon.* **2018**, *68*, 138–145. [[CrossRef](#)] [[PubMed](#)]
130. Pettersson, I.; Ju, W. Design techniques for exploring automotive interaction in the drive towards automation. In Proceedings of the 2017 Conference on Designing Interactive Systems, Edinburgh, UK, 10–14 June 2017; ACM: New York, NY, USA, 2017; pp. 147–160.
131. Lee, J.D.; Moray, N. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* **1992**, *35*, 1243–1270. [[CrossRef](#)] [[PubMed](#)]
132. Lee, J.D.; See, K.A. Trust in automation: Designing for appropriate reliance. *Hum. Factors* **2004**, *46*, 50–80. [[CrossRef](#)]
133. Metz, B.; Landau, A.; Just, M. Frequency of secondary tasks in driving—Results from naturalistic driving data. *Saf. Sci.* **2014**, *68*, 195–203. [[CrossRef](#)]
134. Dingus, T.A.; Klauer, S.G.; Neale, V.L.; Petersen, A.; Lee, S.E.; Sudweeks, J.; Perez, M.A.; Hankey, J.; Ramsey, D.; Gupta, S.; et al. *The 100-Car Naturalistic Driving Study. Phase 2: Results of the 100-Car Field Experiment*; Technical Report; Department of Transportation, National Highway Traffic Safety Administration (NHTSA): Washington, DC, USA, 2006.
135. Gaspar, J.; Carney, C. The Effect of Partial Automation on Driver Attention: A Naturalistic Driving Study. *Hum. Factors* **2019**, *61*, 1261–1276. [[CrossRef](#)]
136. L3 Pilot Consortium. *Deliverable D3.1 From Research Questions to Logging Requirements*; 2018. Available online: <https://cris.vtt.fi/en/publications/from-research-questions-to-logging-requirements-l3pilot-deliverab> (accessed on 15 November 2020).
137. Löcken, A.; Heuten, W.; Boll, S. Enlightening Drivers: A Survey on In-Vehicle Light Displays. In Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ann Arbor, MI, USA, 24–26 October 2016; pp. 97–104.
138. Just, M.A.; Carpenter, P.A. The role of eye-fixation research in cognitive psychology. *Behav. Res. Methods Instrum.* **1976**, *8*, 139–143. [[CrossRef](#)]
139. Kenntner-Mabiala, R.; Kaussner, Y.; Hoffmann, S.; Volk, M. Driving performance of elderly drivers in comparison to middle-aged drivers during a representative, standardized driving test. *Z. Verkehrssicherheit* **2016**, *3*, 73.
140. Naujoks, F.; Wiedemann, K.; Schömig, N.; Jarosch, O.; Gold, C. Expert-based controllability assessment of control transitions from automated to manual driving. *MethodsX* **2018**, *5*, 579–592. [[CrossRef](#)]
141. Jarosch, O.; Bengler, K. Rating of Take-Over Performance in Conditionally Automated Driving Using an Expert-Rating System. In Proceedings of the International Conference on Applied Human Factors and Ergonomics, Orlando, FL, USA, 22–26 July 2018; Springer: Los Angeles, CA, USA, 2018.
142. Pettersson, I.; Frison, A.K.; Lachner, F.; Riener, A.; Nollhage, J. Triangulation in UX Studies: Learning from Experience. In Proceedings of the 2017 ACM Conference Companion Publication, Denver, CO, USA, 6–11 May 2017; pp. 341–344. [[CrossRef](#)]
143. AAM. *Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems*; Alliance of Automobile Manufacturers: Washington, DC, USA, 2006.
144. NHTSA. *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices*; National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT): Washington, DC, USA, 2012.
145. Damböck, D.; Bengler, K. Übernahmezeiten beim hochautomatisierten Fahren. In *5. Tagung Fahrerassistenz; Unfallforschung der Versicherer*: Munich, Germany, 2012.
146. Ghazizadeh, M.; Lee, J.D.; Boyle, L.N. Extending the Technology Acceptance Model to assess automation. *Cogn. Technol. Work* **2012**, *14*, 39–49. [[CrossRef](#)]
147. Eriksson, A.; Stanton, N.A. Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Hum. Factors* **2017**, *59*, 689–705. [[CrossRef](#)] [[PubMed](#)]

148. Frison, A.K.; Wintersberger, P.; Riener, A. Resurrecting the ghost in the shell: A need-centered development approach for optimizing user experience in highly automated vehicles. *Transp. Res. Part Traffic Psychol. Behav.* **2019**, *65*, 439–456. [[CrossRef](#)]
149. Frison, A.K.; Wintersberger, P.; Oberhofer, A.; Riener, A. ATHENA: Supporting UX of Conditionally Automated Driving with Natural Language Reliability Displays. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings (AutomotiveUI'19), Utrecht, The Netherlands, 22–25 September 2019; pp. 187–193. [[CrossRef](#)]
150. Forster, Y.; Frison, A.K.; Wintersberger, P.; Geisel, V.; Hergeth, S.; Riener, A. Where we come from and where we are going: A review of automated driving studies. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings, Utrecht, The Netherlands, 22–25 September 2019; ACM: New York, NY, USA, 2019; pp. 140–145.

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).