Technische Universität München
Lehrstuhl für Ergonomie

Individual Differences in Human-Automation Interaction: A Driver-Centered Perspective on the Introduction of Automated Vehicles

Dipl.-Psych. Univ. Moritz Jürgen Körber

Vollständiger Abdruck der vom Munich Center for Technology in Society der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Philosophie genehmigten Dissertation.

Vorsitzende:	Prof. Dr. Sabine Maasen
Prüfende/-r der Dissertation:	1. Prof. Dr. Klaus Bengler
	2. Prof. Dr. Josef F. Krems
	3. Prof. Dr. Isabell M. Welpe

Die Dissertation wurde am 07.03.2018 bei der Technischen Universität München eingereicht und durch das Munich Center for Technology in Society am 16.07.2018 angenommen.

Acknowledgements

Finishing this thesis would not have been possible without the support and contributions of several people that I would like to thank here. First of all, I would like to thank my supervisor Professor Klaus Bengler for granting me such a great deal of freedom regarding my research. I very much appreciate having had the opportunity to pursue my own research interests and to educate myself during my time at the Chair of Ergonomics.

I would also like to thank Professor Armin Eichinger for inviting me to join the Chair of Ergonomics and for his precious advice at the beginning of my dissertation project. I would furthermore like to thank my colleagues at the Chair of Ergonomics for the fruitful discussions and, more importantly, for the friendly and helpful atmosphere that never made it difficult to come to work. Thanks to you, I have spent a brilliant time at the chair.

Special thanks go to Jonas Radlmayr, Jonas Schmidtler, and Carol-Anne Eberle who took the effort to read my whole thesis and enriched it with their valuable insights and comments. Further thanks go to Jonas Schmidtler and Martin Götze for their constant advise on almost every topic there is and for never hesitating to help or to contribute a witty comment.

Abstract

Following a remarkable increase in traffic efficiency and safety through the introduction of both passive and active safety technologies, automated vehicles are currently being introduced into road traffic with the intention to provide an even higher standard. The introduction of automation into other domains has shown that the anticipated benefits can also be accompanied by unexpected or novel problems. The step from assistance to automation does not simply mean the addition of more of the same but significantly alters the operator's role and responsibilities, and the nature of task demands. Taking this into account, a driver-centered perspective is a key requirement for a successful introduction of automated vehicles.

Driver-centered design requires an understanding of the capabilities and vulnerabilities of both driver and driving automation system, as well as how they vary according to different tasks and situations. This thesis investigates the relevance of individual differences in the driver's interaction with automation with a focus on take-over situations in five studies. A literature review, Article 1, identifies and classifies potential or established individual differences relevant for the interaction with a driving automation system. These individual differences are then positioned within human information processing to deduce their underlying causal mechanisms. The hypotheses about potential individual differences derived from this process are tested in four empirical, confirmatory driving simulator studies. In Article 2, the influence of the driver's age on take-over performance in different traffic densities is investigated. Article 3 studies the ability to multitask sequentially in automated driving, i.e. taking over vehicle control while being engaged in a non-driving-related task. Article 4 investigates the influence of trust in automation on reliance and take-over performance. In the course of this study, a questionnaire to measure trust in automation was developed and evaluated, and is critically discussed. Article 5 investigates the monitoring task in partially automated driving with a focus on the driver's ability to sustain attention and boredom proneness.

The results demonstrate that individual differences among drivers can crucially influence their interaction with automation, which in turn can have a critical impact on safety. An individual's trust in automation influences how much a driver relied on an automated driving system and predicts whether a critical take-over situation will be successfully resolved. The ability to multitask sequentially predicts take-over time during engagement in a non-driving-related task. However, not every driver difference translates into an observable difference in the interaction with automation: Contrary to expectations, there is no evidence for an effect of age on take-over performance and no evidence for a relationship between the ability to sustain attention and the detection of a malfunction. In sum, the studies systematically revealed that individual differences known to be relevant in interaction with automated systems generally can also have a safety-relevant influence on a driver's interaction with a driving automation system. The heterogeneity of the results underlines the importance of taking individual differences into account and highlights the relevance of a driver-centered perspective in automation design.

Zusammenfassung

Die Einführung passiver und aktiver Sicherheitssysteme führte zu einer erheblichen Zunahme der Verkehrseffizienz und -sicherheit. Derzeit werden automatisierte Fahrzeuge mit dem Ziel, diesen Fortschritt noch weiter auszubauen, in den Straßenverkehr eingeführt. Die Einführung automatisierter Systeme in andere Domänen hat gezeigt, dass die antizipierten Vorteile auch mit unvorhergesehenen oder neuartigen Problemen einhergehen können. Der Schritt von Assistenz zu Automation bedeutet nicht lediglich mehr desselben, sondern ändert signifikant die Rolle und die Verantwortlichkeiten des Operateurs sowie die Anforderungen an diesen. Auf Grund dessen ist eine fahrerzentrierte Perspektive ein zentraler Punkt bei der Einführung automatisierter Fahrzeuge.

Fahrerzentrierte Gestaltung setzt ein Verständnis der Fähigkeiten und der Schwachstellen sowie deren Variation in verschiedenen Situationen sowohl von Fahrer als auch Fahrzeugautomation voraus. Aus diesem Grund wurden in dieser Dissertation in fünf Studien individuelle Unterschiede in der Interaktion des Fahrers mit einer Fahrzeugautomation mit einem Fokus auf Übernahmesituationen untersucht. In einer Literaturrecherche, Artikel 1, wurden individuelle Unterschiede, welche für die Interaktion mit einer Fahrzeugautomation relevant sind oder sein könnten, identifiziert und strukturiert. Diese Unterschiede wurde innerhalb des menschlichen Informationsprozesses eingeordnet, um die darunterliegenden kausalen Wirkungsmechanismen zu identifizieren. Die hieraus abgeleiteten Hypothesen wurden in vier empirischen Studien geprüft. In Artikel 2 wurde der Einfluss des Alters auf die Übernahmeleistung bei verschiedenen Verkehrsdichten untersucht. Artikel 3 beschäftigt sich mit der Fähigkeit zu sequentiellem Multitasking beim automatisierten Fahren, d. h. eine Übernahme der Fahrzeugkontrolle während der Beschäftigung mit einer fahrfremden Tätigkeit. In Artikel 4 wurde der Einfluss des Vertrauens in die Automation auf das Blickverhalten und die Übernahmeleistung untersucht. Im Rahmen dieser Studie wurde außerdem ein Fragebogen zu Messung des Vertrauens in Automation entwickelt und evaluiert. In Artikel 5 wurde die Überwachungsaufgabe bei teilautomatisierter Fahrt mit einem Fokus auf die Vigilanzleistung und die Neigung zu Langeweile untersucht.

Die Ergebnisse zeigen, dass individuelle Unterschiede die Interaktion mit einer Automation wesentlich beeinflussen, was sicherheitskritische Folgen nach sich ziehen kann. Das individuelle Vertrauen beeinflusst, wie sehr Fahrer sich auf die Fahrzeugautomation verlassen, und kann vorhersagen, ob eine Übernahmesituation erfolgreich gemeistert wird. Die Fähigkeit für sequentielles Multitasking prädiziert die Übernahmezeit bei gleichzeitiger Beschäftigung mit einer fahrfremden Tätigkeit. Jedoch hatte das Alter in Studie 2 entgegen den Erwartungen keinen Einfluss auf die Übernahmeleistung und es gibt keine Hinweise auf einen Zusammenhang zwischen der Vigilanzleistung und der Entdeckungszeit einer Fehlfunktion. Die Limitationen der Ergebnisse und methodische Herausforderungen in den Untersuchungen werden diskutiert. Die Studien dieser Dissertation haben systematisch gezeigt, dass individuelle Unterschiede, welche in der Interaktion mit automatisierten Systemen relevant sind, auch einen großen Einfluss auf die Interaktion eines Fahrers mit einer Fahrzeugautomation haben können. Die Heterogenität der Ergebnisse unterstreicht die Notwendigkeit, individuelle Unterschiede zu beachten, und verdeutlicht die Relevanz einer fahrerzentrierten Perspektive in der Gestaltung einer Fahrzeugautomation.

Contents

Acronyms

ACC	Adaptive Cruise Control
ADAS	Advanced Driver Assistance Systems
ADS	Automated Driving System
AoI	Area of Interest
BF	Bayes Factor
BPS	Boredom Proneness Scale
CFA	Confirmatory Factor Analysis
CMOS	Complementary Metal–Oxide–Semiconductor
CTT	Classical Test Theory
EFA	Exploratory Factor Analysis
GPS	Global Positioning System
HMI	Human-Machine Interface
KMO	Kaiser-Meyer-Olkin
LIDAR	Light Detection and Ranging
LoA	Level of Automation
NDRT	Non-Driving-Related Tasks
NHST	Null Hypothesis Significance Testing
NHTSA	National Highway Traffic Safety Administration
PEBL	Psychology Experiment Building Language
RADAR	Radio Detection and Ranging
RMSEA	Root Mean Square Error of Approximation
RtI	Request to Intervene
SAE	Society of Automotive Engineers
SEEV	Salience Effort Expectancy Valence
SuRT	Surrogate Reference Task
TiA	Trust in Automation
TLI	Tucker-Lewis Index
TOR	Take-Over Request
TOST	Two One-Sided Tests
TTC	Time to Collision

1 The relevance of a driver-centered perspective for the introduction of automated vehicles

"The question is no longer whether one or another function can be automated but, rather, whether it should be."

(Wiener & Curry, 1980, p. 995)

On May 7th, 2016, Joshua Brown, a 40-year-old man from Ohio (USA), was driving in his Tesla Model S on a state highway in Florida. Suddenly, he collided with a tractor-trailer that was crossing the road in front of his car (Boudette, 2017). But Joshua Brown was not driving himself: The car was operating under its Autopilot system. Neither Autopilot nor Brown stepped on the brakes. Although Autopilot's camera failed to recognize the white truck against the bright sky, the concluding accident investigation "did not identify any defects in the design or performance" of Autopilot (Office of Defects Investigation, 2017, p. 1). Autopilot is a driving automation system, which controls vehicle speed and path. However, the driver has to continuously monitor the driving automation system's behavior as well as the environment and has to be ready to take over vehicle control immediately at any time. Tesla provides information about the driver's responsibilities and the system's limitations at multiple levels, for example in the manual or every time when Autopilot is activated. Therefore, the accident's investigation came to the conclusion that no "incidents in which the systems did not perform as designed" occurred (Office of Defects Investigation, 2017, p. 1). Brown also had enough time to react: With the car's cruise control set at 74 mph, he had at least seven seconds to notice the truck before the crash. But Brown did not brake, steer, or react at all – he was not paying attention to the road. Apparently, he was watching a movie at the time of the collision (Levin & Woolf, 2017).

Parasuraman and Riley (1997) described Brown's behavior almost 20 years earlier as *misuse of automation*. This term describes inappropriate over-reliance on automation when the operator's trust exceeds the automated system's capabilities. Operators then use automation where it should not be used, rely uncritically on automation without considering its limitations, or fall short of monitoring the automated system's decisions or behavior adequately. Brown owned a technology company, was a Tesla enthusiast, and had posted videos on YouTube of him being distracted while driving with Autopilot. In these videos, he marvels that "the car's doing it all itself", taking his hands off the wheel (Brown, 2015). However, data show that Autopilot itself is not an unsafe system. In fact, the frequency of crashes involving Tesla models reportedly declined by about 40 % after its introduction (Office of Defects Investigation, 2017). Still, not everyone trusts automated vehicles as much as Brown did. Although they own the same system, other drivers' videos show them continuously paying attention and even avoiding accidents by overruling Autopilot. It seems that there is a great variability in how people think of and interact with a driving automation system. Indeed, in a recent survey across several industrialized nations, 4 % to 15 % reported no concerns regarding automation technology, whereas up to 44 % indicated that they would monitor the road even if the driving automation system was activated (Schoettle & Sivak, 2014b).

1.1 The introduction of automated driving – expected and unexpected consequences

Several predicted, beneficial outcomes motivate the introduction of automated vehicles to road traffic (Maurer, Gerdes, Lenz, & Winner, 2015; Stanton & Young, 1998; Watzenig & Horn, 2017b): Firstly, driving can be a stressful activity (Matthews, Sparkes, & Bygrave, 1996) and routine drives, such as commuting, have particularly negative effects on the driver's health and mood (Roberts, Hodgson, & Dolan, 2011). Automating this activity would relieve the affected part of the population from this burden and could increase well-being and comfort. Furthermore, driving time could then be used more efficiently since drivers can engage in non-driving-related tasks (NDRTs) such as reading a newspaper (Feldhütter, Gold, Hüger, & Bengler, 2016; Gold, Körber, Lechner, & Bengler, 2016). Besides comfort, automated driving seeks to enhance the mobility of people with medical or age-related mobility constraints, providing them with independence and the requirements for inclusion in economic and social activity (Shergold, Wilson, & Parkhurst, 2016).

Secondly, the German Federal Ministry of Transport and Digital Infrastructure (BMVI, 2015) predicts an increase of 13 % in passenger traffic on German roads by the year 2030. Since the capacity of the traffic system can only be extended to a certain degree, traffic efficiency and degree of capacity utilization must increase. Automated and connected vehicles could help to achieve this goal by reducing speed variability, route planning according to current traffic, and more efficient driving. In addition, fewer traffic jams by means of increased traffic flow could reduce fuel consumption and air pollution as well.

And thirdly, although advancements in passive and active safety technologies have already significantly reduced the number of road accidents (Choi & Ji, 2015), 25 700 road fatalities were still reported in the European Union in 2014 (European Commission, 2015), and this figure increased yet further to 26 000 fatalities in 2015 (European Commission, 2016). This indicates that there is still room for improvement. Human error is the most common cause of road accidents (Singh, 2015), despite the fact that this risk has already been reduced by driver assistance systems (Golias, Yannis, & Antoniou, 2002). It is assumed that automating the driver's tasks will reduce it further still because primary causes of accidents, such as speeding, misjudgment of one's own path, or distraction (Broughton & Markey, 1996; Dingus et al., 2016), may be eliminated.

Currently, legal, ethical, societal, and technical issues of automated driving are being discussed (Maurer et al., 2015). Previous research accompanying the introduction of advanced driver assistance systems (ADAS) has brought to light that to guarantee a successful introduction of a new technology, it is necessary to examine its deployment from a driver-centered perspective as well (Bengler et al., 2014; Körber, Prasch, & Bengler, 2018; Regan, Horberry, & Stevens, 2014). While excellent system performance may be sufficient from a technical standpoint, for it to be accepted and used, a system's functionality must be known, understood, and valued by the driver (Adell, Várhelyi, & Nilsson, 2014; Körber, Prasch et al., 2018; Najm, Stearns, Howarth, Koopmann, & Hitz, 2006). Moreover, besides acceptance, appropriate use is a necessary precondition for the claimed benefits to come into effect. In the process of automation, a function that was previously carried out by a human is fully or partially replaced by a machine (Parasuraman, Sheridan, & Wickens, 2000). Since modern technology is becoming capable of carrying out more and more functions that previously could only be performed by humans, the question arises as to which functions should be automated and to what degree (Parasuraman et al., 2000). The benefits of automation have already been shown in the domain of aviation in the form of cockpit automation (Wickens, Hollands, Banbury, & Parasuraman, 2016). Following a technology-centered approach, automation was seen as the solution to reduce costs and human error. It was implemented whenever it was possible and feasible (Onnasch, 2015; Parasuraman & Riley, 1997). This was based on the assumption that automation can substitute a task formerly executed by a human without any larger impact on the system in which that action or task takes place, predicated on the belief that a complex system consists of a set of independent tasks (Sarter, Woods, & Billings, 1997). However, the limitations of this technology-centered approach as well as initially unanticipated problems have become visible in operational experience (e.g., Three Mile Island incident in 1979) and field research on human-automation interaction. One of the reasons lies in the belief that a human function can be replaced by automation without otherwise affecting the operation of the whole system (Christoffersen & Woods, 2002).

But incidents like Brown's accident strikingly show that technology cannot be considered in isolation from its users. More sophisticated automated systems, such as a driving automation system, represent an increase in autonomy and authority of the machine agent, which in itself is neither good nor bad (Christoffersen & Woods, 2002). Yet, automation does not just replace a human task but rather changes it, sometimes in unanticipated ways. Implementing automation or expanding its role at the same time affects the human's role. It fundamentally alters the nature of the interactions in the system and reshapes the nature of the cognitive demands and responsibilities of the human operator while imposing new coordination demands (Christoffersen & Woods, 2002; Parasuraman et al., 2000; Parasuraman, 2000). Sheridan and Parasuraman (2005) argue that any automation must be designed to perform in conjunction with the human interacting with it rather than expecting the human to adapt to automation. Even advanced automated systems still require human involvement and, for that reason, coordination between human and machine (Sarter et al., 1997). High system performance is, hence, not sufficient. In order to ensure a safe and efficient use, the design of automation must be human-centered (Billings, 1997). That means it must be designed to work cooperatively with the human operator, support human performance, and understanding of the system (Billings, 1997; Christoffersen & Woods, 2002; Riley, 1995). Otherwise, the mentioned anticipated benefits of implementing automation may not become reality or may be offset by human performance costs resulting from maladaptive use of inadequate automation (Parasuraman & Manzey, 2010). Accordingly, recent accounts on automation stress that before considering the design of an automated system "potential system users are identified and characterized for each stage of the system lifecycle" (Wickens, Lee, Liu, & Gordon-Becker, 2014, p. 17). Thus, if we want to avoid the perils of inadequate design of automated vehicles, we first have to have a look at the driver. Let us do that in the next section.

1.2 The relevance of individual differences in research and road traffic safety

While this issue with technology-centered automation became clear in the field of aviation, it may become even more important in the field of automated driving. Contrary to airplane pilots, drivers of automated vehicles will have limited training and will have passed through minimum screening and examination. Therefore, the population of drivers is more heterogeneous and variable with respect to individual differences, such as in cognitive performance, impairments, age, or affinity for technology (Creaser & Fitch, 2015). Those individual differences are an important point to consider for multiple reasons.

Firstly, individual differences are often seen as annoying noise in the data that spoil statistical significance tests. Here is Cronbach (1957) portraying this view:

> Individual differences have been an annoyance rather than a challenge to the experimenter. His goal is to control behavior, and variation within treatments is proof that he has not succeeded. Individual variation is cast into that outer darkness known as 'error variance'. . . . your goal in the experimental tradition is to get those embarrassing differential variables out of sight. (p. 674)

Yet, individual differences may contain information that is crucial for drawing valid conclusions. Many theories provide a good fit regarding the average performance of a group of participants (Parasuraman, 2011). Yet, these models may not necessarily apply to some or even many individuals within a group. Even creating two groups (e.g., slow and fast responders) with adjusted model parameters still may not be sufficient to create a coherent description. For example, Parasuraman (2009) shows that the decrement of vigilance (a state or degree to detect infrequent and randomly occurring target stimuli among frequent non-targets over a prolonged period of time) in a group can be well fitted by an exponential function. However, in the referenced study, only 40 % of the participants showed a vigilance decrement over time whereas the remainder showed stable or variable detection rates. Theory also suggests that driving performance is impaired if the driver is engaged in another task (e.g., talking on a cell phone) at the same time. Yet, Watson and Strayer (2010) report that 2.5 % of their sample showed no decline driving performance while being engaged in a difficult secondary task. It is well known that if a study is supposed to produce representative results valid for a population of interest, a representative sample instead of a specific group of participants (e.g., test drivers) has to be drawn. Yet, even if this is considered, a focus on group means may provide a misleading image. A mean curve does not allow to make statements regarding the corresponding individuals because it could result from any of an infinite collection of populations of individual curves (Estes, 1956; Sidman, 1952). Estes (1956) warns that "the uncritical use of mean curves even for such purposes as determining the effect of an experimental treatment . . . is attended by considerable risk" (p. 134). Furthermore, group means are also inappropriate if the goal is to draw a conclusion that is valid for the whole population. In this case, it is crucial to consider the boundaries of a distribution (Körber & Bengler, 2014). For example, Sohn and Stepleman (1998) recommend using the 85th or 99th percentile instead of the mean reaction time when making a suggestion for a safe headway distance.

Secondly, the relevance of individual differences for road traffic safety also becomes apparent by the fact that governmental institutions exist that assess an individual's fitness to drive (Bukasa & Utzelmann, 2009). Their relevance can also be directly observed in road accident statistics: Accident-proneness, a conjectured pre-disposition to be involved in accidents, describes the observation that some individuals are disproportionally often involved in recurring crashes caused by human error (Das, Sun, Wang, & Leboeuf, 2015; Visser, Pijl, Stolk, Neeleman, & Rosmalen, 2007). Following this notion, some drivers simply seem to be more likely in general to be involved in an accident than others. But other, more specific relevant individual differences have been identified as well: In their analysis of US crash data, Stutts, Reinfurt, Staplin, and Rodgman (2001) identified distraction as the cause of an accident in 8.3 % of the cases. Engaging in complex secondary tasks while driving manually increases an individual's near-crash/crash risk by three (Klauer, Dingus, Neale, Sudweeks, & Ramsey, 2006). The involvement in such an accident can be traced back to individual differences among the drivers: Engagement in distracted driving is associated with individual risk perception (Rupp, Gentzler, & Smither, 2016), which is in turn associated with a higher crash risk (Ivers et al., 2009; Oltedal & Rundmo, 2006). Sitkin and Pablo (1992) specify in their risk model that risky behavior, besides contextual factors, is determined by the individual characteristics risk preference, risk perception, and risk propensity. Wilde (1982), in the same manner, argues that risk behavior is highly dependent on the individual because it is mainly determined by an individual's striving to establish a subjective risk homeostasis rather than by objective risk alone. Depending on the perceived objective risk, individual risk behavior is adapted so that ultimately a homeostasis between the subjectively perceived risk level and the preferred risk level is achieved. Young, in particular male drivers, exhibit more risky driving behavior and perceive situations as less risky (DeJoy, 1992; Evans & Wasielewski, 1983; Finn & Bragg, 1986). Drivers who are impulsive or who are sensation seekers engage more frequently in risky driving manners such as speeding and violating safe driving laws (Arthur, Barrett, & Alexander, 1991; Burns & Wilde, 1995; Schwebel et al., 2007). Thus, these mentioned individual differences in the end partly determine involvement in crashes caused by distraction.

Closely related to risky behavior are individual differences in driving experience. Novice drivers generally overestimate their own driving skills and also accept more risks while driving (Deery, 1999). The ability to avoid collisions increases with a driver's experience because drivers acquire a more efficient search strategy (Koustanaï, Boloix, van Elslande, & Bastien, 2008) and a more effective gaze behavior (Underwood, Chapman, Brocklehurst, Underwood, & Crundall, 2003). Drivers also get better at allocating their attention during secondary task engagement with more driving experience (O'Brien, Klauer, Ehsani, & Simons-Morton, 2016). As a result, driving experience and a driver's expectations are the main determining factors of brake reaction times (Green, 2000; Horswill & McKenna, 2004). Consequently, crash risk decreases with each subsequent year of driving experience, independent of age at the time of licensure (Maycock, Lockwood, & Lester, 1991; Twisk & Stacey, 2007). Besides this, expertise with a system promotes complacent behavior (Singh, Molloy, Mouloua, Deaton, & Parasuraman, 1998), influences risk perception (Hoedemaeker & Brookhuis, 1998; Rajaonah, Tricot, Anceaux, & Millot, 2008) and trust (Rudin-Brown & Parker, 2004).

Lastly, road accidents can also be traced back to differences in cognitive functioning and ability. Individuals who performed better in cognitive tasks show a more effective eye movement strategy while driving and exhibit better driving performance (Mackenzie & Harris, 2017). Kahneman, Ben-Ishai, and Lotan (1973) as well as Mihal and Barrett (1976) accordingly reported that cognitive abilities such as selective attention are correlated with road accidents. The reported frequency of everyday slips and errors is positively correlated with driving error rates (Allahyari et al., 2008) and the number of accidents (Larson & Merritt, 1991). The ability to multitask, i.e. to be engaged in a secondary task while driving, also has an individual difference component itself (Morgan, D'Mello, Abbott et al., 2013; Watson & Strayer, 2010) and is determined by working memory performance as well as other cognitive abilities (Bühner, König, Pick, & Krumm, 2006; Tijerina, Parmer, & Goodman, 1998). In general, the design of interfaces is guided by considerations on workload, trying to avoid both underload and overload. However, individual workload is largely determined by individual working memory performance (Ahmed et al., 2014; König, Bühner, & Mürling, 2005; Parasuraman & Jiang, 2012; Parasuraman, Sheridan, & Wickens, 2008).

The studies highlight that individual differences have serious implications for road traffic. Experience with automation in other domains has shown that a human-centered design approach is necessary to overcome the pitfalls that loom when humans interact with automated systems. Given the current introduction of automated vehicles to road traffic, the question arises if a human/driver-centered design approach also is a crucial stringer to ensure a successful introduction. More precisely, are individual differences also relevant in the interaction with driving automation systems? Motivated by this question, this thesis aims to systematically investigate if individual differences matter in the interaction with driving automation systems. The remainder of this thesis is structured as follows: First, an introduction to automation and automated driving is given in Section 2. Next, individual differences in information processing as a causal mechanism for individual differences in the interaction with automation are discussed in Section 3. Based on this, the five research articles that constitute this thesis are presented in Sections 4 to 8. Theory, background, and potential causal mechanisms are discussed in this introduction and are completed by a summary of each article. A discussion of the articles' results follows in Section 9, including limitations and objectives for future research. The content of Sections 4 to 9 is based on or extracted from the respective article. At the end, a final conclusion on the scientific contribution and the achieved progress is drawn and practical implications in form of recommendations are given. Appendix A provides additional complementary articles and methods for this thesis. The content of Appendix A is based on or extracted from the respective article.

2 Automated driving

"Any cars that are being made that do not have full autonomy will have negative value. It will be like owning a horse. You will only be owning it for sentimental reasons."

Elon Musk (as cited in Thompson, 2015)

2.1 Automation: Definition, types, and levels

Automation can generally be defined as "the full or partial replacement of a function previously carried out by the human operator" (Parasuraman et al., 2000, p. 287). Several frameworks have been proposed to classify automated systems regarding the distribution of functions between human and machine (Manzey, Reichenbach, & Onnasch, 2012). Automated systems can differ in type and complexity, from merely acquiring information to fully autonomous execution of an action. In their model, Parasuraman et al. (2000) classify automated systems regarding two aspects: The *type of an automated* system describes what is being automated, i.e. the classes of functions that can be carried out by the automated system. The type can be directly mapped to the stages of the human information processing (Figure 1), which will be described in more detail in Section 3. Depending on the type of automation, a task of information processing is transferred to the automated system that is then executed by it from now on. The second aspect represents the *level of automation* (LoA) within these four types, i.e. how much of this function is automated (Figure 2; Manzey, 2012). The LoA describes the allocation of the task, from no allocation (manual control) to completely autonomous execution by automation (Sheridan, 1992). Both dimensions together have been subsumed by Wickens, Li, Santamaria, Sebok, and Sarter (2010) in the term *degree of automation*. Thereby, automation is not all-or-none but varies across types and levels on a continuum from manual to fully automated.

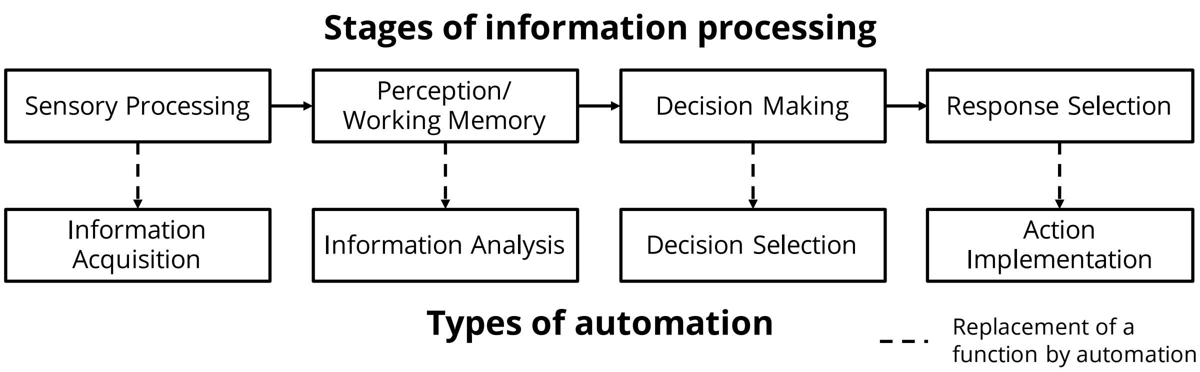

Figure 1. Depending on the type of automation, a different stage of information processing is replaced by the automated system.

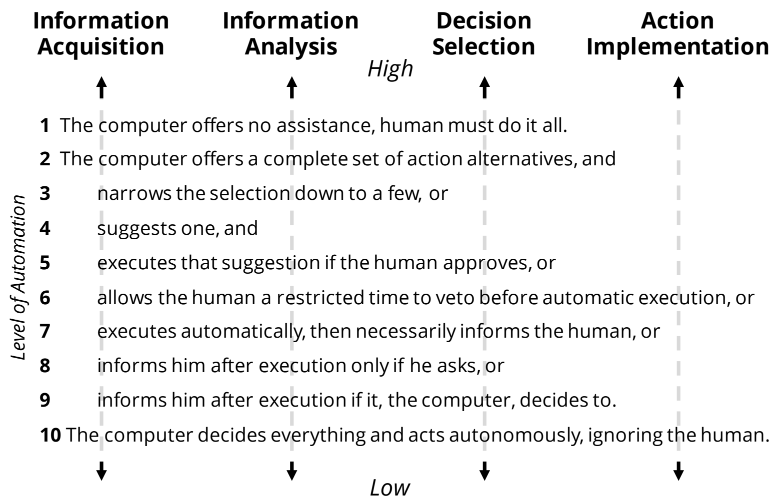

Figure 2. The model of different types of automation by Parasuraman et al. (2000) combined with the levels of automation by Sheridan (1992).

2.2 Driving automation systems and levels of driving automation

Automated driving also represents the transfer of a task – the primary driving task – from human to automation. Bubb (2015) classifies the *primary driving task* into three hierarchical tasks, *navigation* (selection of a driving route), *maneuvering* (planning the precise trajectory, lane changes), and *stabilization* (vehicle control necessary to execute the planned maneuvers). Additionally, the driver has to perform *secondary tasks*, which are not directly related to vehicle guidance but have to be performed for reasons associated with traffic or the environment, such as activating the windshield wipers. Optionally, the driver can also engage in *tertiary tasks* which serve to increase comfort or entertainment, such as turning on the radio. The dynamic interaction between driver, vehicle, and environment can be described by a closed driver-vehicle feedback loop (Figure 3).

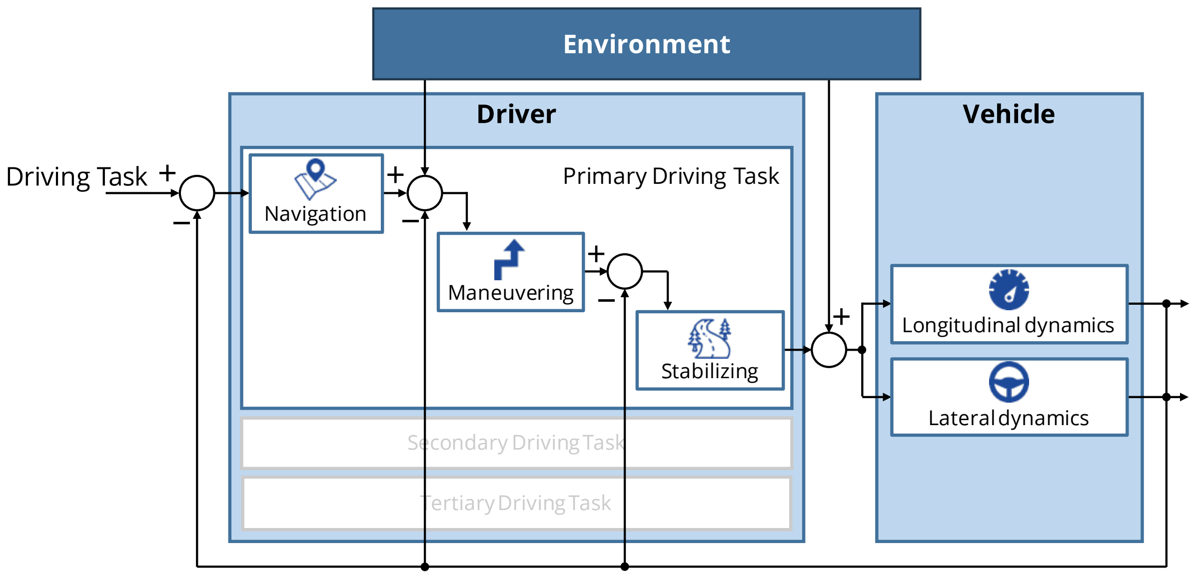

Figure 3. The driver-vehicle feedback loop; adapted from Bubb, Bengler, Grünen, and Vollrath (2015).

The driving task represents the set point, which underlies disturbances from the environment (roadway arrangement, other traffic, weather etc.). Departures from this set point value are recognized by the driver/controller, who controls the effector's/vehicle's longitudinal and lateral dynamics in order to re-establish the set point.

To describe how the primary driving task is transferred to a driving automation system, six levels of automation are used (Figure 5; SAE International, 2016) instead of the taxonomy developed by Parasuraman et al. (2000). While Level 0 corresponds to manual driving without any assistance, Level 1 represents assisted driving where the human is still executing vehicle control assisted by ADAS. The introduction of a driving automation system removes the driver from this driver-vehicle-feedback loop and transfers the primary driving task to the driving automation system (to varying degrees; Figure 4). Brown's Autopilot by Tesla, which was mentioned in the introduction, represents a Level 2 partial automation. This level represents the first real transfer of vehicle control from the driver to a driving automation system. Here, the function that is transferred is longitudinal (accelerating, braking) and lateral (steering) vehicle control. Nonetheless, the driver has to constantly monitor the environment and supervise the driving automation system. He has to be ready to take over vehicle control immediately. In Level 3, conditionally automated driving, the task of monitoring the environment is also transferred to an automated driving system (a Level 3, 4, or 5 driving automation system; ADS). The driver is completely removed from the feedback loop and can engage in non-driving-related activities such as reading a newspaper. The driver merely acts as a fallback level and has to take over vehicle control with a certain lead time if requested to do so by the ADS due to a system limit.

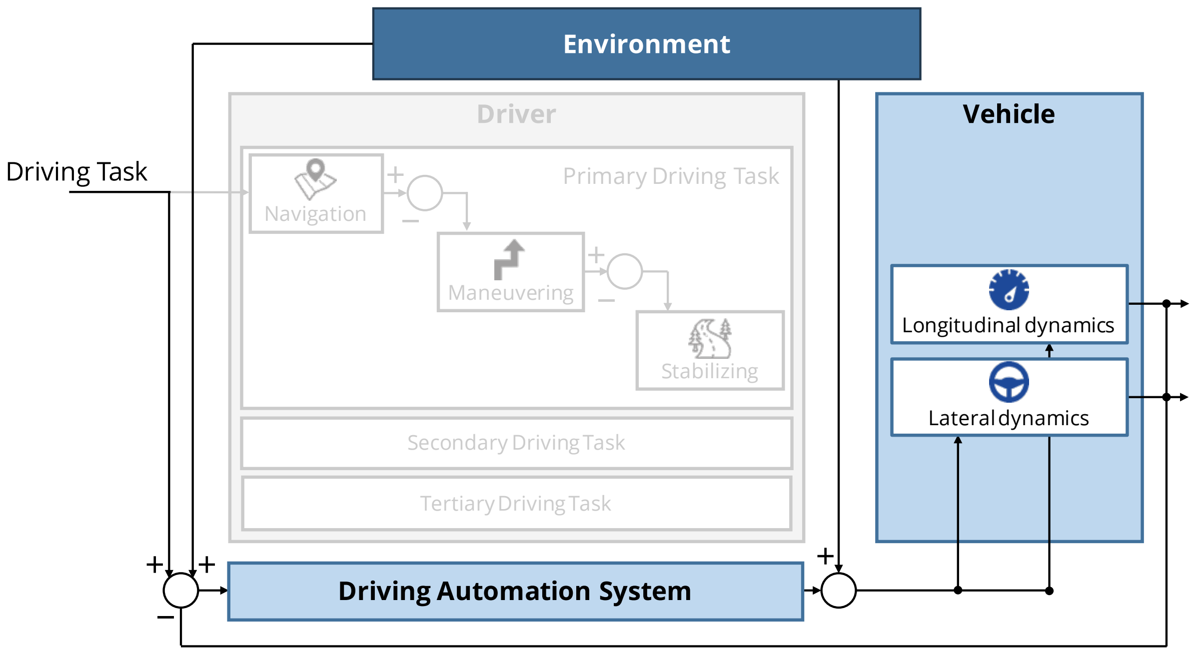

Figure 4. Automated driving represents the transfer of the primary driving task to a driving automation system.

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/ Deceleration	*Monitoring* of Driving Environment	Fallback Performance of *Dynamic Driving Task*	System Capability *(Driving Modes)*
***Human driver* monitors the driving environment**						
0	No Automation	the full-time performance by the *human driver* of all aspects of the *dynamic driving task*, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	Driver Assistance	the *driving mode*-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	Human driver and system	Human driver	Human driver	Some driving modes
2	Partial Automation	the *driving mode*-specific execution by one or more driver assistance systems of both steering and acceleration/ deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	**System**	Human driver	Human driver	Some driving modes
***Automated driving system* ("system") monitors the driving environment**						
3	Conditional Automation	the *driving mode*-specific performance by an *automated driving system* of all aspects of the dynamic driving task with the expectation that the *human driver* will respond appropriately to a *request to intervene*	System	**System**	Human driver	Some driving modes
4	High Automation	the *driving mode*-specific performance by an automated driving system of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a *request to intervene*	System	System	**System**	Some driving modes
5	Full Automation	the full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	System	**All driving modes**

Figure 5. The six levels of automation according to SAE International (2016).

Table 1 explains the role and responsibilities of the driver and ADS in greater detail. Other taxonomies referred to this level as highly automated driving (Gasser, 2012) or Limited Self-Driving Automation (NHTSA Level 3; National Highway Traffic Safety Administration, 2013). The fallback role is also transferred to the ADS in Level 4. While Level 4 is not available in every situation or condition, Level 5 represents an unconditional completely self-driving system. The studies of this thesis mainly focus on Level 3 automation.

The introduction of driving automation cannot be considered in isolation from the driver, especially since the driver will still play a relevant role even in higher automation levels (Christoffersen & Woods, 2002; Gasser & Schmidt, 2017). Automated driving is distinct from driver assistance both in a quantitative and qualitative manner (Gasser & Schmidt, 2017). The transfer of vehicle control represents a change in the driver's role and the resulting physical and cognitive demands. Levels 2 to 5 represent an increase in the number of functions the driving automation system is now responsible for as well as the situations in which this driving mode is possible. While the numeration of the levels suggests a quantitative, gradual increase in the degree of automation, they, in fact, each represent qualitatively different stages. Automation does not simply replace an independent task previously executed by a human, but rather creates a whole new human-machine system. Increasing the LoA profoundly changes the driver's role, the coordination between driver and system, and the type and extent of the cognitive demands (Parasuraman et al., 2000; Parasuraman, 2000). For example, engagement in an NDRT occupies working memory and promotes a shift of attention away from the traffic scene (Baumann, Rösler, & Krems, 2007).

Table 1
The role of user and driving automation system in conditionally automated driving (Level 3) according to SAE International (2016)

Role of User	Role of Automated Driving System
Driver (while the ADS is not engaged): – Verifies operational readiness of the ADS-equipped vehicle – Determines when engagement of ADS is appropriate – Becomes the dynamic driving task fallback-ready user when the ADS is engaged Dynamic driving task fallback-ready user (while the ADS is engaged): – Is receptive to a request to intervene and responds by performing dynamic driving task fallback in a timely manner – Is receptive to dynamic driving task performance-relevant system failures in vehicle systems and, upon occurrence, performs dynamic driving task fallback in a timely manner – Determines whether and how to achieve a minimal risk condition – Becomes the driver upon requesting disengagement of the ADS	ADS (while not engaged): – Permits engagement only within its operational design domain ADS (while engaged): – Performs the entire dynamic driving task – Determines whether operational design domain limits are about to be exceeded and, if so, issues a timely request to intervene to the dynamic driving task fallback-ready user – Determines whether there is a dynamic driving task performance-relevant system failure of the ADS and, if so, issues a timely request to intervene to the dynamic driving task fallback-ready user – Disengages an appropriate time after issuing a request to intervene – Disengages immediately upon driver request

Consequently, being engaged in a secondary task extends the reaction time (Petermann-Stock, Hackenberg, Muhr, & Mergl, 2013), even if it is merely a cognitive task (Radlmayr, Gold, Lorenz, Farid, & Bengler, 2014). Each LoA represents a human-machine system on its own, with its own specific demands, responsibilities, coordination, and pitfalls. What statement might be true for Level 2 does not have to be valid for Level 3 and vice versa. Research conducted in different areas has shown that the more support an automated system provides, the better the performance if it works flawlessly. But at the same time, failure entails a steeper drop in performance (Onnasch, Wickens, Li, & Manzey, 2014). Thereby, considering, safety, ease of use, costs, and liability, it is still open for discussion "which system functions should be automated and to what extent" (Parasuraman et al., 2000, p. 286).

This approach is not the only possible concept of the distribution of control. In cooperative (guidance and) control, the driver and a driving automation system are cooperatively controlling the vehicle in varying shares (Flemisch, Bengler, Bubb, Winner, & Bruder, 2014). This paradigm may be realized at different levels of the driving task. In shared control, human and automation may work on the vehicle control task at the same time. Both are continuously sharing the control authority and jointly determine the input. The underlying idea is to keep the driver in the direct manual control loop while still providing continuous support (Mulder, Abbink, & Boer, 2012; Petermeijer, Abbink, & de Winter, 2015). It may also be possible to cooperate on the maneuver guidance level by delegating subtasks to automation in a hierarchical or adaptive manner. *Conduct-by-Wire* (Franz, Kauer, Geyer, & Hakuli, 2016) is such a maneuver-based realization of cooperative guidance. This paradigm represents a static and hierarchical distribution of control between the driver as a maneuvering commander and the driving automation system executing the stabilization. Albeit drivers have no role in stabilization anymore, they are still in the control loop by executing the guidance task. The maneuver interface also provides a permanent fallback in case of system limits.

The concept *H-Mode* is inspired by the cooperation between rider and horse (Flemisch et al., 2014). H-Mode transfers this cooperation to vehicle guidance and represents an approach with an emphasis on a haptic-multimodal coupling between driving automation (horse) and the driver (rider). The quintessence of this approach is vehicle control in form of a multi-modal combination of the driving automation system's intent and driver input via an active interface, and fluid transitions between two levels of automation: *tight rein* (assisted) and *loose rein* (highly automated; Flemisch et al., 2014; Kienle, Damböck, Kelsch, Flemisch, & Bengler, 2009). Later a third mode, *secured rein*, which takes the driver completely out of the loop, was added. Both driver and automation are able to initiate transitions between these three levels of automation, either by buttons or fluidly by tightening/loosening or even discontinuing the grip on the active interface.

2.3 Underlying technology of automated vehicles

In the following, a brief introduction into the underlying technology of automated driving is given. An account in greater detail is given, for example, in Eskandarian (2012), Winner, Hakuli, Lotz, and Singer (2016), or Watzenig and Horn (2017a). The technologies that are required for realizing automated driving on highways can be allocated to three main branches: perception of the environment, vehicle localization, and driving strategy (Kämpchen, Aeberhard, Ardelt, & Rauch, 2012).

To move safely in traffic, a vehicle has to know the exact location and attributes of all objects around it. Thus, perception of the environment is necessary to reliably detect all objects and relevant traffic. The employed sensory systems can be seen as a modest extension of the sensors that are currently integrated and used for various driver assistance applications. Car manufacturers rely on different configurations that are constituted of mainly four different types of sensors:

– RADAR: Radar sensors emit pulses of electromagnetic waves and enable detection of vehicles and obstacles, including their speed and direction of motion, in the front and the rear of the automated vehicle. Depending on the system, long-range systems (77 GHz) generally cover a distance up to 250 m with a small spread (ca. 18°), whereas short-range systems (24 GHz band) cover about 70 m at a higher spread of about 90–130° (Kirschbaum, 2015; Stevenson, 2011).

- LIDAR and laser scanner: These sensors use ultraviolet, visible or infrared light pulses for object detection and, by sensor fusion, provide a gapless surround view of the vehicle's environment. Limited in range, they are responsible for the perception of the environment at lower distances than Radar sensors but provide a more detailed image. Laser scanners functionally extend LIDAR systems by providing a dynamic viewing angle by rotating sensors (Levinson & Thrun, 2010).
- Ultrasonic: The ultrasonic sensors are implemented on the side of an automated vehicle and provide (redundant) detection of close objects at a short range. The sensors use sound waves (around 48 kHz) to calculate the distance between the sensor and an object, with a detection range of 25 to 400 cm (Paulweber, 2017).
- Mono or stereo camera: Image sensors (e.g., CMOS) in cameras deliver a detailed front/rear view (horizontal field of view of about 45° to 90°) and allow a reliable classification of objects at a range of 50 to 100 m. Thereby, the vehicle does not only know that an obstacle but also what obstacle is present. Furthermore, they detect lane markings for localization, evaluate road quality, and read traffic signs and signals. When using multiple cameras (stereo cameras), depth of field can be included. Infrared systems can be used for night vision (Punke, Menzel, Werthessen, Stache, & Höpfl, 2016).

To ensure system robustness, the sensors are implemented following the principle of redundancy, i.e. no area is only covered by a single sensor or only covered by utilization of a single measurement modality. Sensor data is then processed, organized, and fused (Kämpchen et al., 2012). Figure 6 shows an exemplary sensor configuration.

Accurate trajectory planning and vehicle control require a precise location of the vehicle on the road and highly accurate digital maps that provide information on the current and upcoming course (e.g., the curvature of the curves). For this purpose, the global location determination of the vehicle on a high precision multi-lane map by GPS and odometry is used. This information is combined with a sensor road model, created by the sensor data from cameras and laser scanners, to enable a centimeter-precise localization of the vehicle (Rauch, Aeberhard, Ardelt, & Kämpchen, 2012).

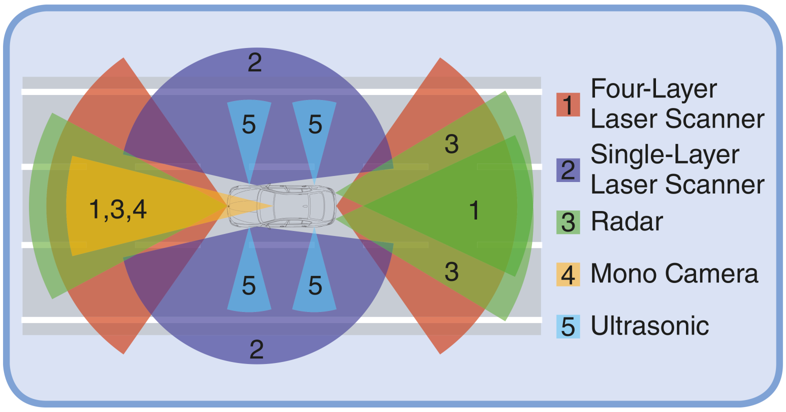

Figure 6. A possible sensor configuration of an automated vehicle illustrated by Aeberhard et al. (2015).

Building on environmental perception and vehicle localization, a driving strategy derives the best possible course of action that moves the vehicle safely and comfortably based on the current traffic situation and the driver's goals. Appropriate trajectories are generated by artificial intelligence, which then sends them to a robust and real-time vehicle controller for execution (Aeberhard et al., 2015). For example, if a slower vehicle is detected ahead, the system analyses the surrounding environment and traffic to check if an overtake maneuver is possible and feasible. The system then executes the lane change, carried out fully automatic, while constantly monitoring traffic and the upcoming road (Kämpchen et al., 2012).

2.4 Take-over requests

Conditionally automated driving (Level 3) still requires a driver as a fallback level. The ADS may request the driver to take over vehicle control within a certain lead time. This take-over request (TOR), also called request to intervene (RtI; Marberger et al., 2018), may be triggered, for example, because of missing map data, leaving the operational design domain, conditions inhibiting sensor perception, sensor failure, or unpredictable or too complex situational factors (Gold, 2016). Such system limits are detected by the ADS, which emits the request to monitor or to take over manual vehicle control. Such a TOR might be signaled by a form of visual, auditory, vibro-tactile, or multi-modal stimulus (van den Beukel, van der Voort, & Eger, 2016; Petermeijer, Cieler, & de Winter, 2017). This signal is emitted at a certain lead time, the time budget to react.

Several authors (Gold & Bengler, 2014; Petermeijer, de Winter, & Bengler, 2016; Zeeb, Buchner, & Schrauf, 2015) have identified multiple stages of the take-over process, which are more or less based on the information processing model by Wickens et al. (2016). The response to a TOR consists of partly automatic and partly controlled processing (Schneider & Shiffrin, 1977). Firstly, the driver senses the TOR, then perceives it, and, in an automatic response, orients his attention towards the road ahead (Zeeb et al., 2015). This is a highly practiced response that is rapid and without capacity limitation, because there is no conscious perception, processing, or decision making involved – the signal processing wanders straight from sensation to motor reaction (Green, 2000). Subsequently, a driver interprets the current situation regarding status, attributes, and dynamics of the relevant elements such as the position of other vehicles or the upcoming course (Endsley, 1995). The perceived information is processed and interpreted while it is being aligned with the driver's working memory (e.g., a comparison with the status at the last gaze to the driving scene) and long-term memory (e.g., expectations, traffic rules, schemata). Parallel to information processing, the driver returns to a driving position (e.g., feet on pedals, hands on steering wheel), which is again a highly trained automatic response. The result of this process is the selection of a response, once more in cooperation with long-term and working memory. This response, for example evasive steering, is then executed as a motor reaction (Gold & Bengler, 2014). Contrary to the first automatic response, these steps represent conscious controlled processing, which requires thought, is slower, serial, effortful, and capacity-limited (Schneider & Detweiler, 1988). However, automatic and controlled reactions do not represent a strict dichotomy but rather a graded continuum because the response in some situations can be an automatic process as well.

For example, drivers often unconsciously hit the brakes if they sense flashed brake lights ahead (Green, 2000). Figure 7 illustrates the particular phases of a take-over process identified in Petermeijer et al. (2016).

Several metrics can be derived to evaluate this process, which may be more or less arbitrarily classified into performance metrics and quality metrics. Performance is defined as "any activity or collection of responses that leads to a result" (VandenBos, 2015, p. 778), whereas the result here can be seen as a successful take-over, while the definition of success depends on the take-over scenario. For example, if the take-over situation requires evading an obstacle, a situation is successfully solved if no collision occurred (Körber, Gold, Lechner, & Bengler, 2016). Thus, the simplest, yet most relevant metric of take-over performance is if the situation has been solved successfully. However, this metric also conveys the lowest amount of information; drivers who did not react at all are classified the same way as drivers who barely could not avoid a collision. Chronometric measures represent a measurement on ratio scale level and are a detailed representation of the degree to which the situation has been solved successfully. Because of these properties, studies to investigate the influence of certain characteristics of the take-over situation primarily rely on these measures (Gold et al., 2016; Gold & Bengler, 2014; Gold, Damböck, Lorenz, & Bengler, 2013; Körber, Gold et al., 2016). Analogously to the chronometric experiments on information processing in cognitive psychology, the duration of the several stages of the take-over process, such as gaze reaction time or hands-on time, can be assessed individually for a detailed evaluation (Gold, 2016). The time between the emittance of the TOR and the first controlled reaction of the driver is called take-over time (TOT) and represents the main metric since it subsumes the speed in all other previous stages of the take-over process. Certainly, this definition is incomplete without defining the term conscious reaction. Different variables and thresholds for conscious reaction result in different values for the take-over time for the exact same data (Gold & Bengler, 2014; Payre, Cestac, Dang, Vienne, & Delhomme, 2017; Zeeb et al., 2015).

Secondly, the quality of a take-over can be evaluated. Quality refers to the characteristic, the essential character, or nature of something (VandenBos, 2015). Take-over quality is a metric that subsumes how the take-over has been executed and, indirectly, to what degree it was safe. The take-over scenario determines how a successful take-over looks like and what metrics are meaningful to measure take-over quality in this scenario. For evasive maneuvers, as investigated in this thesis, time-to-collision (TTC) is a common metric for the criticality of a situation (Winner et al., 2016) and is thereby also a metric for take-over quality. The TTC is defined as the current

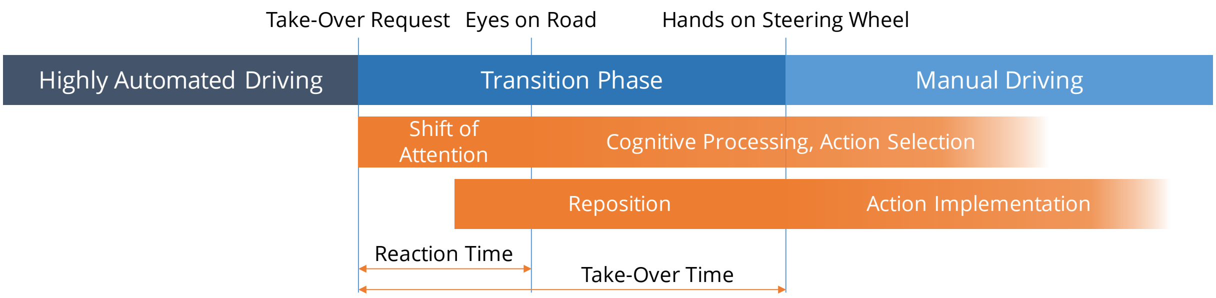

Figure 7. Phases of a take-over process according to Petermeijer et al. (2016).

remaining time until a collision with an object on the assumption of constant speed and direction for both (while corrections exist for other cases; Winner et al., 2016). The minimum value represents the most informative descriptive statistic, which marks the point of the highest criticality occurred. Besides the TTC, vehicle dynamics, such as longitudinal and lateral vehicle accelerations, describe how the take-over was performed. Again, the minimum and maximum value are the most informative descriptive statistic.

All metrics have to be interpreted in the context of the whole maneuver. Absolute values may be misleading: An uncritical take-over does not necessarily imply low accelerations and high accelerations do not necessarily indicate a critical take-over maneuver. For example, braking instead of no braking does not directly mean that the take-over was, in fact, more critical. A participant that did not brake and changed lanes without checking the mirrors may act riskier despite exhibiting lower longitudinal accelerations. In the same manner, the potential presence of a speed-accuracy trade off has to be taken into account (Hong & Williamson, 2008): A participant may take his time to choose an appropriate reaction. In this case, the participant's take-over time is higher while in the end the reaction might be better or safer. That being said, it has to be taken into account that take-over performance and take-over quality are not completely independent dimensions but are correlated with strength varying on the take-over scenario. Moreover, absolute take-over quality, i.e. independent of the situation, differs from relative take-over quality, dependent on the possibilities of a situation. In sum, a holistic evaluation of the take-over process instead of mere reliance on numbers is recommended. A more detailed account of the process of a take-over can be found in Gold (2016).

For a successful introduction of automated vehicles, reliable system performance has to be ensured from the technological side. Yet, while vehicle control is transferred to a driving automation system in automated driving, drivers are not completely released from their responsibilities in Level 2 to 3 automated driving. It thus has to be ensured that drivers can interact with a driving automation system in a safe manner as well. Section 1.2 already highlighted the importance of individual differences in driving behavior and in the interaction with automation. Taking this into account, this thesis aims to investigate the influence of individual differences on the interaction with driving automation with a focus on take-over situations.

3 Information processing and behavior – causal mechanisms of individual differences

"Science walks forward on two feet, namely theory and experiment . . . .
Sometimes it is one foot that is put forward first, sometimes the other,
but continuous progress is only made by the use of both"

Robert A. Millikan, Nobel Lecture 1924 (as cited in Gigerenzer, 2009)

How do the aforementioned individual differences in the interaction with automation arise? How can it be that high trust in automation leads to a collision? To investigate these research questions, it is necessary to embed the findings in the current state of research and to create a scientific theory that includes an underlying causal mechanism. Establishing a causal relationship allows performing forward causal inference (What will happen if X?) as well as reverse causal inference (What causes Y? Gelman, 2011). Experiments themselves do not lead to causal explanations but only to causal descriptions (Shadish, Cook, & Campbell, 2002). In other words, they describe the influence of the systematic variation of an independent variable on a dependent variable but do not offer explanations why and under which conditions a certain relationship exists. The identification of a causal mechanism, in contrast, opens the possibility to estimate the generalizability of results. It may also help to explain why a replication of an experiment may have failed and helps to lower the probability of being fooled by spurious relationships without an underlying causal relation. According to Popper (1959), for a theory to be scientific, it has to make falsifiable predictions. Scientific progress is then the process of proposing falsifiable theories and testing their predictions with observations. Observational evidence can refute the theory or, if they are in line with the predictions, corroborate it (Dienes, 2008; Howson & Urbach, 2006). Indeed, while cognitive ergonomics mainly deals with descriptive analysis, Wickens et al. (2016) consequently distinguish engineering psychology from cognitive ergonomics because only the former has a strong and necessary basis in theory. Thus, the first step of this thesis in the investigation of individual differences in the interaction with automation was to establish a suitable theoretical model in order to explain and to consolidate the findings. After this, the therefrom derived hypotheses were tested by confirmatory studies presented in this thesis.

3.1 Models of information processing

Depending on the LoA, manual and automated driving can be conceptualized as a task that primarily consists of information acquisition and transforming this information into an adequate reaction (Abendroth & Bruder, 2016). It is essential to take information processing into account when investigating the interaction between humans and automation because in almost all such interactions the operator has to perceive and process information, react based on the result, and then must evaluate the feedback from that action, i.e. its effect on the environment (Wickens & Carswell, 2006). Thus, to pinpoint individual differences, first the relationship between information processing and response needs to be investigated. Information processing models may function as a causal explanation. They could explain why something happens (Why does trust differ?) but also predict what will happen (What is the behavioral outcome of a difference in trust?).

The paradigms of human information processing characterize humans as a system that receives input from the environment, processes that input, and then returns a reaction as output to the environment (Proctor & van Zandt, 2008). The models are typically elaborated versions of this three-stage model consisting of perception, cognition, and reaction. Their construction follows an inductive approach by the aggregation of experimental studies of human performance that provide their empirical base (Proctor & van Zandt, 2008). Researchers identified the stages by aggregating the results of many single experiments that are based on chronometric methods with reaction times as the primary measure of information processing (Lachman, Lachman, & Butterfield, 1979). By manipulating the characteristics of tasks that should influence a certain assumed processing stage, the influence of this manipulation can be measured as the change in the resulting reaction time (Donders, 1969; Proctor & Vu, 2009; Sanders, 2013; Sternberg, 1969). Further support for distinct stages comes from neurobiological studies that identified distinct brain regions that correspond in their activity to the different proclaimed stages (Wickens & Carswell, 2006).

Information processing represents a deeply complex process (Pashler, Yantis, Medin, Gallistel, & Wixted, 2004). In order to understand and work with this process, an abstraction in form of a model is necessary. A model functions as an abstract, simplified representation of a process that strives to approximate reality. Wickens et al. (2016) provide a basic model of human information processing (Figure 8), which gives a useful overview of the relevant stages of information processing. Let us go through the model stage by stage: In the proclaimed flow of information, stimuli are first sensed or attended by the sensory system with the quality and quantity of the information (and of all other following processing stages) being dependent on the properties of the sensory system in question. The information is then stored in the sensory memory for a very brief time. It is then processed for the first time in the perceptual stage where it is consciously

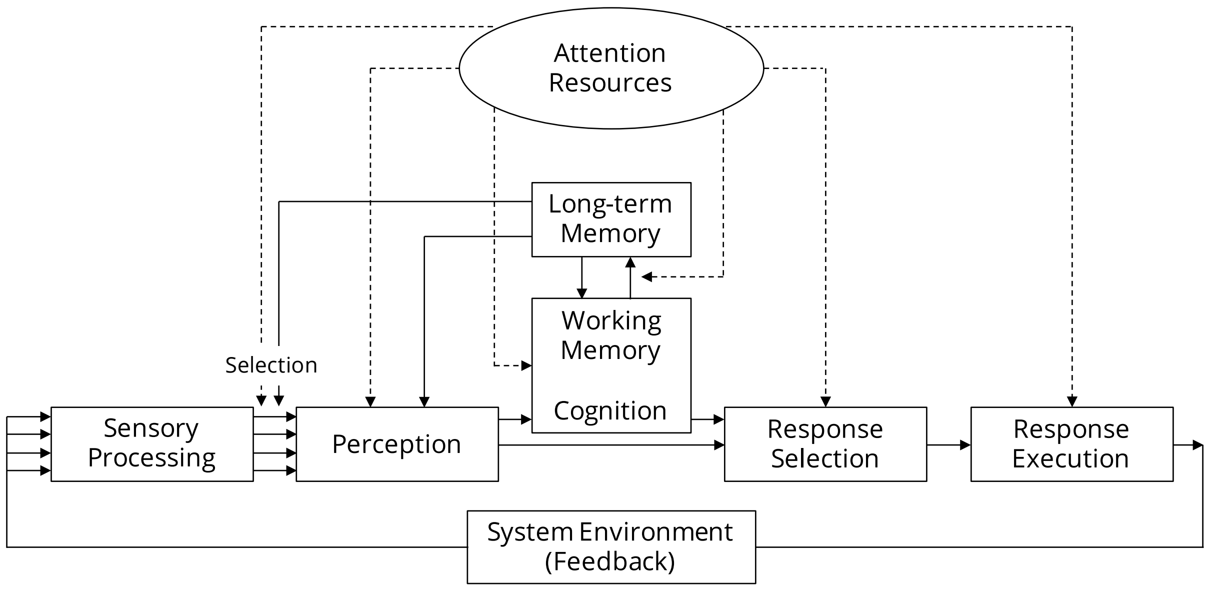

Figure 8. The information processing model of Wickens et al. (2016).

perceived and interpreted. In the information processing stage, the decision to act or not to act is made. This decision is significantly determined by the individual. The type of action can be classified into three levels (skill, rule, or knowledge-based behavior; Figure 9; Rasmussen, 1983). *Skill-based* behavior comprises automatic sensumotoric reactions with low processing time without conscious top-down regulation (e.g., adapting vehicle speed to a headway vehicle). *Rule-based* behavior requires more cognitive resources and is a conscious action led by stored rules or heuristics that are activated based on the cues provided by the situation at hand (e.g., interacting with other road users). *Knowledge-based* behavior represents the most resource-straining behavior that comes into play in novel situations where no stored rules apply (e.g., navigation in an unfamiliar environment). The current situation is analyzed in detail and, depending on personal preferences, a plan of action is constructed.

The whole process of perception and response selection is moderated by an operator's attention and memory. Following Baddeley's (2007) multicomponent working memory model, memory can be, in a simplified manner, structured into a short-term working memory and a long-term memory (besides the already mentioned sensory store; Sternberg, Sternberg, & Mio, 2012). Working memory functions as a temporary, attention-demanding store, where novel, as well as currently-demanded information, is stored for a short time. Corresponding illustrating metaphors for working memory are a workbench of consciousness or a computer's random-access memory. Guided by the central executive that controls and regulates the cognitive processes, humans use their working memory to examine, evaluate, transform, and compare mental representations of information, which are then used for action or stored in long-term memory (Wickens & Carswell, 2006). Working memory is responsible for processing and manipulating the perceived information and choosing an adequate action. Long-term memory in turn stores different variants of memory (such as procedural, semantic, episodic) that can be retrieved if needed in the current operation (Atkinson & Shiffrin, 1968). It stores rules, past experiences, models, and facts that influence detection, perception of information, and, hence, how we see the world.

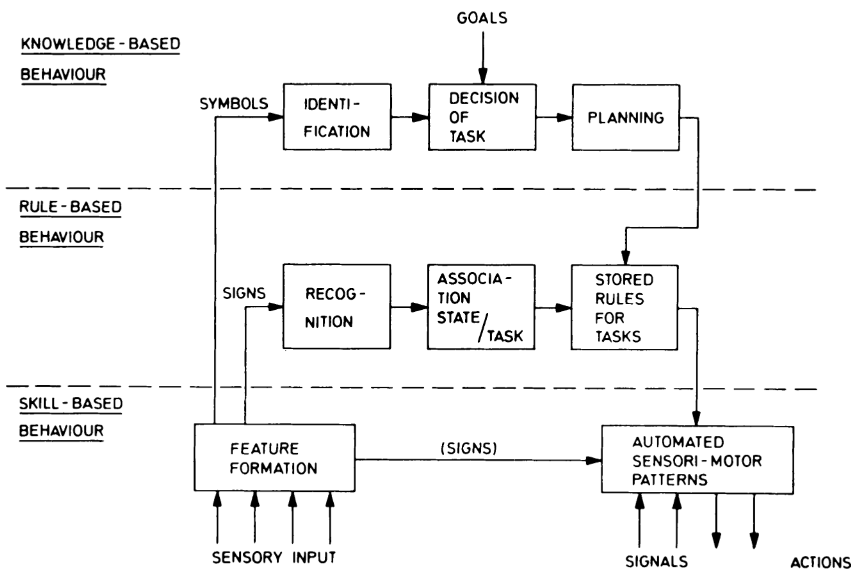

Figure 9. Three types of human behavior according to Rasmussen (1983).

Attention, the second moderator, acts as the fuel to carry out all of the mentioned processes. This fuel or mental resource can, in function of a filter, be selectively allocated to stimuli, information, or the processing stages and, thereby, selects what information will be processed in what level of detail and what tasks can be performed simultaneously (Wickens & Carswell, 2006).

The last stage represents the execution of the response, which has been selected based on the results of the prior stages. Evoked feedback is then sensed. This creates a closed-loop cycle, where the feedback from a system to a response is in the end perceived and evaluated again, mimicking closed-loop models of control engineering (Jagacinski & Flach, 2003).

The model certainly simplifies the process to provide a helpful and functional overview. For example, it does not detail the underlying operations in the different processing stages and subsumes very complex processes, such as perception, in a single box. It is thereby not exhaustive of all the relevant cognitive processes (Pretz & Sternberg, 2005; Sternberg et al., 2012). Moreover, while the whole process is moderated by attention, there is no detailed elaboration on whether the attentional capacity is limited by a single resource (Kahneman, 1973) or multiple resources (Wickens, 2002). Due to this simplification, the model can be consulted to explain a variety of effects, which is its strength but also its weakness. The applicability of the model should not be over-generalized. For example, the model is not suitable for highly practiced automated actions such as manually shifting a gear while driving, which are processed effortless and subconsciously without an involvement of the working memory (Schneider & Shiffrin, 1977). However, "all models are wrong but some are useful" (Box, 1979, p. 202): Contrary to rather isolated paradigms in cognitive psychology produced by laboratory research, the model strives to be applied in naturalistic contexts and gives an accurate overview of the underlying processes. Differences in the observable outcome variable, i.e. the response execution, can be traced back to differences in a prior processing stage. Individual differences in states, traits, and dispositions can be found at each processing stage, causing different responses from these individuals.

Sanders (1983) and Luczak (1975) developed very similar models, supporting the construct validity of the model of Wickens et al. (2016). Following a combination of stage-wise information processing (as in Wickens et al., 2016) and resource models (Kahneman, 1973; Luczak, 1998), Abendroth and Bruder (2016) provide a model that transfers information processing to the driving context and describes the interaction between driver and vehicle in great detail. However, this model is less parsimonious than the model of Wickens et al. (2016) and provides an abundance of parameters that are in parts not specified in detail. This makes the model very complex, i.e. it is able to fit various or arbitrary data patterns. Such a model is difficult to falsify and it is difficult to determine how much of a good fit can be attributed to verisimilitude compared to the model's general ability to fit any arbitrary data (Preacher, Zhang, Kim, & Mels, 2013). For this reason, this thesis relies on the information processing model of Wickens et al. (2016) to explain individual differences among drivers in their interaction with automation by tracing them back to differences in the stages of the information processing model.

3.2 Positioning individual differences in information processing

Rohmert (1984) as well as Bubb (1992) highlight that the outcome of a human-machine-interaction not only depends on factors that are common for each individual but it also on the individual's reaction to these factors, which is determined by the characteristics of this individual. Following this notion, an individual characteristic component is implemented in almost every model that is relevant for human-machine interaction, such as human error (Gründl, 2005; Reason, 2000), situation awareness (Endsley, 1995), or mental workload (de Waard, 1996), as well as in several driver behavior models (Irmscher & Ehmann, 2004; Markkula, Benderius, Wolff, & Wahde, 2012; Yang, Rakheja, & Stiharu, 2001).

Consequently, Abendroth and Bruder (2016) added an individual characteristics component to their mentioned model. The individual characteristic parameter interacts with the information processing at every stage and, thereby, moderates the outcome. However, in this thesis, another taxonomy of individual characteristics based on the domains of biopsychology and personal psychology building on Körber and Bengler (2014) is proposed. Firstly, following the state-trait-model known from differential psychology (Amelang & Schmidt-Atzert, 2006) and latent state-trait theory (Kelava & Schermelleh-Engel, 2012), an individual characteristic is classified as either state or trait, depending on whether the characteristic remains stable over time and is consistent in different situations (trait) or not (state). However, the boundaries are fuzzy yet meaningful (Allen & Potkay, 1981; Chaplin, John, & Goldberg, 1988), forming rather two ends of a continuum than a rigid dichotomy. States are considered temporary, brief, and mainly determined by external circumstances. In the driving context, for example, the whole research domain of driver state monitoring focuses on the detection of a driver's current state, such as current drowsiness (Schmidt, Braunagel, Stolzmann, & Karrer-Gauss, 2016). Contrarily, a (phenotypic) trait is defined as a stable attribute resulting from an interaction of a hereditary predisposition and environmental influence (VandenBos, 2015) and are either directly observable (e.g., height) or non-manifest latent constructs (e.g., personality traits, such as conscientiousness, or attitudes). Those characteristics vary in their stability as well as in their dependence on a genetic disposition. For example, abilities, as listed by Abendroth and Bruder (2016), are considered by VandenBos (2015) as an innate existing competence to perform an act, whereas skills (as listed in the model) are acquired through training.

The causes of these aforementioned individual differences can be located in the information processing model of Wickens et al. (2016). For example, experienced drivers differ from novice drivers in their allocation of attention (Crundall, Underwood, & Chapman, 1999; Underwood et al., 2003), the moderator variable in the model. For them, the driving task has become an automated process, which leads to differences in the capacity utilization of the model's working memory component (Schneider & Shiffrin, 1977). They also have more strategies internalized and can recall them with less effort (Chase & Simon, 1973; Krings et al., 2000), which makes them differ in the response selection and the decision-making stage (Pashler & Baylis, 1991; Pretz, 2008; Wiggins, Stevens, Howard, Henley, & O'Hare, 2002). Lastly, training improves the response execution (Logan, 1992; McPherson, 1999). Differences in the ability to drive distracted can be traced back to differences in the working memory component (König et al., 2005). Drowsiness deteriorates attention (Goel, Rao, Durmer, & Dinges, 2009) or, for example, the decision-making process (Lobb & Stern, 2009), attention to and processing of risk-relevant cues as well as cognitive control (Damasio, 1996; Rao, Korczykowski, Pluta, Hoang, & Detre, 2008). The majority of these findings can already be traced back to differences in the neuroanatomy and neurophysiology (e.g., Anderson, Vicki, & Jacobs, 2014; Haines & Schenk, 2015). The recently emerged field of Neuroergonomics, "the study of the human brain in relation to performance at work and other everyday settings" (Parasuraman, 2011, p. 181), looks into the underlying causes on the very basic level of gene expression and molecules (Parasuraman & Rizzo, 2008). For example, Parasuraman, de Visser, Lin, and Greenwood (2012) successfully investigated the neurobiology of automation bias, the tendency to erroneously follow incorrect advice or information provided by the automated system. The extent to which individuals exhibit automation bias varies with their capacity of working memory and the efficiency of their executive functions. Both are highly heritable and under dopaminergic and noradrenergic control in the prefrontal cortex.

In sum, these findings answer why individual differences among drivers in manual or automated driving exist. We can trace the differences in driving back to differences in the drivers' information processing, which in turn correspond to differences in neuroanatomy and neurophysiology.

3.3 Individual differences in human-automation interaction and in automated driving

The aforementioned study by Parasuraman et al. (2012) shows that individual differences also exist in the domain of automation. There is more evidence: Especially in dual-task situations, some operators tend towards complacency, i.e. they shift their attention away from an automated task to a concurrent task, resulting in monitoring failures (Parasuraman et al., 2008). The tendency to this maladaptive attentional strategy appears to be determined by the operator's personality and is a stable personality trait (Singh, Molloy, & Parasuraman, 1993a). Furthermore, participants who prefer manual operation over automated operation tend to lapse into complacency more quickly (Maehigashi, Miwa, Terai, Kojima, & Morita, 2012). Performance in general benefits from working with a functioning automation but suffers when automation is erroneous, particularly at higher levels of automation (Onnasch et al., 2014). Rovira, Pak, and McLaughlin (2016) showed that the extent of the costs and benefits depend on individual differences in working memory ability. In their study, performance of participants with low working memory ability suffered more severely when automation failed at a high LoA than those with higher working memory. Lower working memory capacity was also related to higher trust in automation in their study. de Visser, Shaw, Mohamed-Ameen, and Parasuraman (2010) reported similar results on the role of working memory in an automated unmanned aerial vehicle task.

Previous research has also identified individual differences in automated driving. As described in Section 2.2, depending on the LoA, implementing a driving automation system replaces the driver's active role of manual driving with either a passive role as a monitor or as a fallback level in case of a system limit. Such a situation of task underload and monotony has been shown to induce hypovigilance, passive fatigue, and drowsiness (Körber, Cingel, Zimmermann, & Bengler, 2015).

Accordingly, Feldhütter, Hecht, Kalb, and Bengler (2018) reported an increase in drowsiness during a prolonged automated drive, however, the observed onset and extent of this drowsiness increment were highly variable among the participants. Error bars in studies on drowsiness caused by an automated drive also similarly show large individual differences regarding the onset or extent (Körber, Cingel et al., 2015; Schmidt et al., 2009; Schmidt et al., 2016). Moreover, Merritt and Ilgen (2008) found that besides actual objective characteristics of an automated system, such as its reliability, the individual subjective perception of its characteristics has a significant impact on trust. The subjective perception is in turn determined by the operator's personality, or, more precisely, the general propensity to trust. In this study, the participants' general propensity to trust also determined the magnitude of the effect of automation failures on trust in automation. Individual trust in automation can also explain differences in NDRT engagement since it determines how much participants monitor the environment during the engagement with an NDRT (Hergeth, Lorenz, Vilimek, & Krems, 2016). Consequently, the heterogeneity in trust in the population also shows up in surveys on automated driving across several industrialized nations: Less than 15 % reported absolutely no concerns regarding automation technology, whereas up to 33 % would refuse to ride in an automated vehicle (Schoettle & Sivak, 2014a, 2014b, 2015, 2016).

Motivated by these findings, the first aim of the research process of this thesis was to identify and structure the abovementioned findings in a literature review on potential and known individual differences regarding automated driving. This was realized in Article 1.

4 Article 1: "Potential individual differences regarding automation effects in automated driving"

Körber, M., & Bengler, K. (2014). Potential individual differences regarding automation effects in automated driving. In C. S. G. González, C. C. Ordóñez, & H. Fardoun (Eds.), *Interacción 2014: Proceedings of the XV International Conference on Human Computer Interaction* (pp. 152–158). New York, NY, USA: ACM.

Motivated by the need to investigate the influence of individual differences in automated driving, this article investigates the characteristics of automated driving and lists individual differences that could potentially influence human performance in interaction with driving automation.

At the beginning, the characteristics of the interaction with Level 2 and Level 3 driving automation, such as a passive monitoring role in Level 2, have been defined. Then, findings from the general domain of automation, driver assistance systems, engineering psychology, general psychology, and research on human performance have been screened for their relevance regarding these characteristics. The literature review uses a slightly deviating version of the aforementioned state-trait model and structures the found individual differences into dispositional factors, stable personality traits and behavior patterns, current operator state, attitudes, and demographic factors. Since automated driving is a novel technology, most of these individual differences have not yet been empirically investigated and, thus, only represent potential individual differences. The aim of this dissertation thesis was to empirically investigate these potential individual differences in automated driving.

The results of this literature review laid the foundations for a confirmatory research agenda on the empirical investigation of individual differences. Theories on a possible individual difference have been proposed, hypotheses have been deducted from them, and their predictions have been tested on the basis of observations in four confirmatory studies (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). This led to the following four articles of this thesis.

5 Article 2: "The influence of age on the take-over of vehicle control in highly automated driving"

Körber, M., Gold, C., Lechner, D., & Bengler, K. (2016). The influence of age on the take-over of vehicle control in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *39*, 19–32.

5.1 A closer look at the process of aging

Industrialized nations are currently witnessing a steadily growing proportion of elderly people in the driving population – the global population older than 65 years is predicted to double from 7 % to 14 % by 2040 (Cauley, 2012). Automated vehicles are hypothesized as an option to enhance the mobility of elderly adults (Reimer, 2014; Shergold et al., 2016). Thus, elderly drivers will represent a relevant and even increasing proportion of users of automated vehicles.

Aging influences information processing in the model of Wickens et al. (2016) on every stage and component, for example executive functions, working memory, attention, or speed of information processing (Anstey, Wood, Lord, & Walker, 2005; Bryan & Luszcz, 2000; Der & Deary, 2006; Miller, Taylor-Piliae, & Insel, 2016; Salthouse, 2009). For example, the sensory perception of elderly adults is impaired (Haegerstrom-Portnoy, Schneck, & Brabyn, 1999), there is a decline in the perception of hazards (Horswill et al., 2008), and in the speed of information processing (Salthouse, 1991; Verhaeghen & Salthouse, 1997). Working memory function, as well as executive functions, decline with age (Bryan & Luszcz, 2000; Salthouse & Babcock, 1991). Elderly adults have a decreased ability to divide attention (Ponds, Brouwer, & Van Wolffelaar, 1988) and their dual-task performance is worse (Baldwin & Schieber, 1995; Hartley & Little, 1999). In addition, they suffer from a decline in their ability to maintain and to select task sets, and in task switching (Kray, Eber, & Lindenberger, 2004; Kray & Lindenberger, 2000; Mayr, 2001). Consequently, their reaction times are higher (Der & Deary, 2006) and elderly drivers exhibit higher brake reaction times and hazard perception response times (Horswill et al., 2008; Warshawsky-Livne & Shinar, 2002). Anstey et al. (2005) provide an overview of age-related changes relevant to safe driving and report associations of varying strength between the decline in performance in cognitive tests with driving outcome measures. Deficiencies in attentional, perceptual, cognitive, and psychomotor abilities predict unsafe driving incidents and involvement in motor vehicle crashes (Emerson et al., 2012; Mathias & Lucas, 2009; McKnight & McKnight, 1999). While the reflection of age-related decline in accident statistics is still under discussion (Schlag, 2013), elderly drivers are overrepresented in crashes in complex situations such as intersections (Braitman, Kirley, Ferguson, & Chaudhary, 2007; Clarke, Ward, Bartle, & Truman, 2010).

Automated driving still requires a driver in certain situations (see Section 2.2). Because reaction times (Der & Deary, 2006), hazard perception response time (Horswill et al., 2008), processing speed (Salthouse, 1991), and task switching performance (Kray & Lindenberger, 2000) deteriorate with age, it may be possible that elderly drivers could have difficulties to take over vehicle control in time in critical situations. Moreover, automated driving allows the engagement with an NDRT in certain modes. Elderly adults have a lowered ability to shift attention flexibly between two tasks, to adhere to a prioritized focus (Siu, Chou, Mayr, van Donkelaar, & Woollacott, 2008), and a decreased ability to divide attention (Ponds et al., 1988). In addition, they show deficits in the ability to keep up relevant situational information (Salthouse, 1991). Correspondingly, elderly people take longer to resume a task following an interruption (Monk, Boehm-Davis, & Trafton, 2004).

However, albeit every aging individual experiences this decline, there is a high variability in the decline's rate and intensity (Hultsch, MacDonald, & Dixon, 2002), which further increases with age (Morse, 1993). Hertzog, Kramer, Wilson, and Lindenberger (2008) describe this decline as a "zone of possible functioning" (p. 1) whose limits are determined by individual endowment and age-related constraints. An individual's position in this zone depends on the individual engagement in beneficial intellectual, physical, and social activities. Given the right conditions, satisfactory cognitive functioning can be maintained even at high age. Besides this, performance in everyday naturalistic tasks does not solely depend on general cognitive ability but also on task-specific knowledge and expertise (Masunaga & Horn, 2001). Deterioration of cognitive performance is thereby not exclusively a function of age but also of compensatory adaptation, experience-related changes, and acquisition of expertise. Altogether, aging does not automatically cause a deterioration in driving performance but its impact strongly depends on the specific situation and the specific driver.

The impact of age-related decline on driving performance has already been investigated in manual driving (Devlin, McGillivray, Charlton, Lowndes, & Etienne, 2012; Horberry, Anderson, Regan, Triggs, & Brown, 2006). Yet, the novelty of the technology of automated driving and the qualitative differences in the driver's role and tasks between manual and automated driving (described in Section 2.2) create the need to investigate the interaction of elderly drivers with driving automation. As already mentioned, elderly drivers will represent an increasingly important user segment. Safety has to be ensured for every potential user, not just for an average user (Körber & Bengler, 2014). This is especially important given the mentioned variability in age-related cognitive decline. In this study, we investigated the influence of age on take-over situations in conditionally automated driving.

5.2 Summary of Article 2

Seventy-two participants divided into two age groups, consisting of 36 elderly drivers (≥ 60 years) and 36 younger drivers (≤ 28 years), took part in this study. The conditions of the take-over situation were manipulated by adding a verbal NDRT (20 questions task) and by variation of traffic density. The participants encountered three situations with either no, medium, or high traffic density in which they had to take over vehicle control to evade an obstacle on their lane. We found no significant difference between the take-over times of elderly drivers and younger drivers. However, the age groups differed in their modus operandi: The elderly drivers maintained a higher TTC, a higher proportion in this group hit the brakes during the take-over, and braking was stronger. While the engagement in an NDRT had no effect, there was negative influence of a higher traffic density on take-over time and take-over quality independent of age. Both younger and older drivers showed a learning effect between the first and the last take-over situation in form of a decrease in take-over time, an increase in minimum TTC, and a decrease in maximum lateral acceleration.

The results highlight that despite the mentioned assumed decline in cognitive performance, elderly participants did not significantly differ from young participants in their take-over times. Both age groups adapted to the experience of the three take-over situations in the same way and to the same extent. While the study provides initial evidence on the role of age in take-over situations, it is difficult to generalize these results to the whole population of elderly drivers. First, cognitive decline is highly variable in its degree and speed. Both depend on factors such as personal lifestyle, individual experiences, and genetics (Deary et al., 2009; Hultsch et al., 2002; Morse, 1993). Variance in these factors leads to variance in the cognitive ability. Second, not aging in general but a decline in particular abilities may cause an age-related increase in crash risk (Hakamies-Blomqvist, 1998). For these reasons, elderly drivers at the lower end of the zone of possible cognitive functioning (Hertzog et al., 2008) might solve the take-over situations less successfully, despite being the same age.

6 Article 3: “Prediction of take-over time in highly automated driving by two psychometric tests”

Körber, M., Weißgerber, T., Kalb, L., Blaschke, C., & Farid, M. (2015). Prediction of take-over time in highly automated driving by two psychometric tests. *Dyna*, *82*(193), 195–201.

6.1 Multitasking revisited

Wickens et al. (2016) describe multitasking not only as dividing attention between information channels but also as dividing attention between tasks. The best-known account on multitasking may be Wickens’s (2002) theory on multiplicity, illustrated as a three-dimensional cubic structure. In this theory, he claims that the extent to which two tasks demand separate resources along four dimensions (stage, code, modality, visual location) determines the decrement in task performance in comparison to single-task execution. This account revised Kahneman’s (1973) model of a single pool of undifferentiated capacity.

However, this is only one component of his multiple resource theory. Task interference is also determined by the difficulty, i.e. the total resource demand of a task (Figure 10). Even if they use similar resources, two very easy simultaneously executed tasks can still be solved successfully. In the end, which task suffers, how interruptions are managed, or what information is processed is under the regulation of the central executive system, which is responsible for selective attention, inhibition, or shifting between tasks in a top-down manner (Sternberg et al., 2012; Wickens & McCarley, 2008). This system is also responsible for shifting attention between multiple activities, even if no dual-task situation is present. Salvucci (2013) construes multitasking as a continuum between concurrent multitasking and sequential multitasking. Concurrent multitasking embodies the act of doing two tasks at the same time. As the intervals between the task switches grow, multitasking more and more embodies a sequential task paradigm. It follows that multitasking does not only comprise simultaneously engaging in two tasks but also situations in which users switch

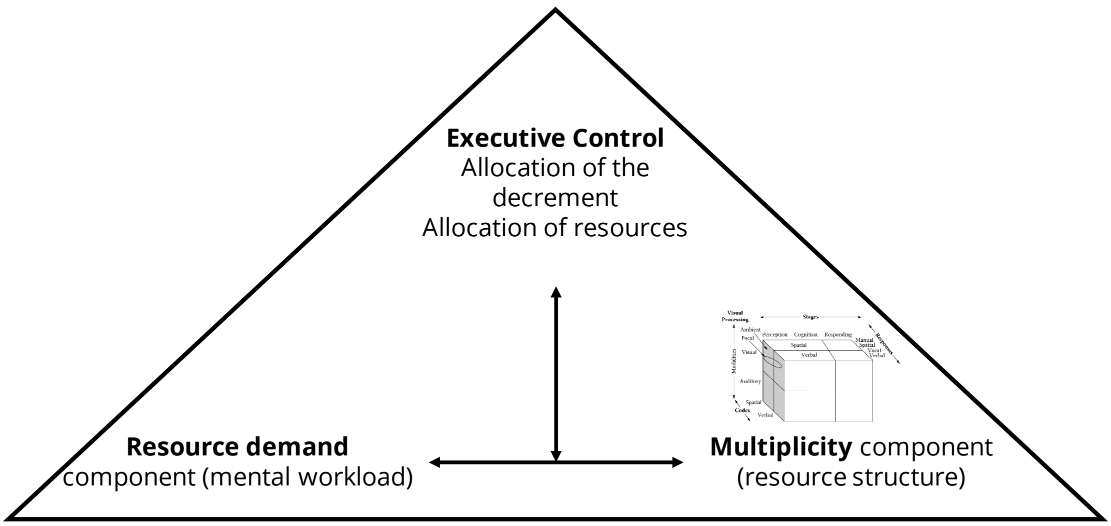

Figure 10. The architecture of the multiple resource theory of Wickens (2002); adapted from Wickens et al. (2016).

between tasks after minutes or abandon one task entirely for a certain amount of time, either motivated by task interruption or task management. Such conditions are also prevalent in conditionally automated driving. The complete transfer of the dynamic driving task to driving automation allows a driver to engage in an NDRT. For example, a driver may look up from his newspaper to check why the vehicle performs a lane change. Task-relevant information, details of the visual scene, and similar information is retained in the driver's working memory – and will be updated by his next gaze on road. As a result, the driver has to switch between two tasks if a TOR occurs. Switching between tasks results in switch costs: An individual's responses are slower and more error-prone directly after a task switch (Monsell, 2003).

Other problems that may arise from replacing an operator with an automated system in the feedback loop are subsumed under the term *out-of-the-loop problem* (Endsley & Kiris, 1995)[1]. Engagement in an NDRT promotes the driver to orient his attention away from the diving scene and occupies working memory (Baumann et al., 2007). The rehearsal and retrieval of task-specific information in working memory have been proposed as a bottleneck in multitasking. This process consumes a so-called *problem state resource*, which is used to maintain the mental representations that are necessary for executing a task (Borst, Taatgen, & van Rijn, 2010). Interference arises when state-related information of the driving task (e.g., location of other vehicles) and of the NDRT are stored, manipulated, and retrieved from working memory at the same time. For example, active engagement in a driving task leads to less road scanning than passively monitoring (Mackenzie & Harris, 2015). This, in turn, may result in a loss of situation awareness, "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1988, p. 97). According to Endsley and Kiris (1995), operators who have lost situation awareness may be slower to diagnose problems, require more time to re-orient themselves to the relevant parameters, and take longer to resume manual control. If the driver is engaged in an NDRT when a TOR is issued, he has to reallocate his attention to the driving scene, regain situation awareness, and react appropriately. Performance in this situation is a function of reaction time, the storage of task-relevant information in working memory, and the ability to switch tasks.

Multitasking performance is more than simply the sum of two single-task performances. The performance decrement that arises out of a dual-task situation is correlated with single-task performance, but the single-task performance cannot fully explain the variance in dual-task performances. This unexplained variance reflects the component that is unique to dual-task situations (Wickens et al., 2016). Once more, the question arises why individuals differ in their ability to multitask and, as before, the information processing model can be applied. A first cause may lie in the cognition stage and the following stages of response selection and execution. Individual differences in time sharing between two tasks can be traced back to differences in

[1] Removing the driver from the driver-vehicle loop may induce an out-of-the-loop state defined as "a driver state of readiness in which the driver is not able to immediately intervene in the feedback loop comprised of controller and vehicle. In this state, the driver does not have up-to-date knowledge of the parameters that are relevant for the controlling task, e.g. his own speed, position, or a headway vehicle. He is also not able to predict the situation insofar as to create a time window for himself that is long enough to react to events in a manner that is safe for road traffic" Körber, Weißgerber, Kalb, Blaschke, and Farid (2015, p. 196).

expertise in one or both of the tasks: Experts exhibit automated task execution and thus reduced effort to execute an action (Fisk, Ackerman, & Schneider, 1987). Wickens et al. (2016) further suggest that experts exhibit more efficient visual scanning and greater attentional flexibility. Mackenzie and Harris (2017) explain the differences between novices and experts in scanning behavior (Underwood, Crundall, & Chapman, 2002) by two mechanisms. Novice drivers may not be aware of potentially hazardous areas and, consequently, do not know where to look. But driving is also a less automated task for novices, which takes up the majority of their attentional resources. For example, working memory load has been found to elicit spatial gaze concentration on the road center (Recarte & Nunes, 2003) and impairs object detection (Törnros & Bolling, 2006). Hence, novices lack the attentional resources to orient visual attention to areas relevant for safe driving, which makes them vulnerable to performance costs in task switching situations.

Beyond that, stable individuals differences in working memory capacity, also a component of the information processing model, exist (Engle, 2002). Working memory is responsible for the ability to maintain information in an active, quickly retrievable state, and to control attention, both of which are important for multitasking. Working memory moreover supports individuals to switch from one task to another by storing information related to a task that they are not currently executing, and by controlling attention (König et al., 2005). This conceptualization of working memory is accommodated in the multidimensional model of working memory by Oberauer, Süß, Wilhelm, and Wittman (2003). In this model, the three dimensions of working memory are *storage in the context of processing*, *coordination*, and *supervision* (Bühner et al., 2006). Under high task load, these three functions break down. Accordingly, working memory was the most important predictor of multitasking performance in addition to attentional performance and fluid intelligence in a study by König et al. (2005) and in a study by Bühner et al. (2006). Age-related differences in working memory are also seen as the cause of a decline in task-switching with increasing age (Kramer, Hahn, & Gopher, 1999; Kray, Li, & Lindenberger, 2002).

Findings that investigate these hypothesized individual differences in multitasking in applied settings are sparse. Participants in the study of Wood, Hartley, Furley, and Wilson (2016) with lower working memory capacity performed poorer in a dual task, reported more instances of inattention during driving and exhibited poorer hazard perception performance under dual task conditions. Morgan, D'Mello, Adams et al. (2013) found working memory to be a significant predictor of multitasking ability in a flight simulation. Recently, Mackenzie and Harris (2017) found that participants who performed better in two multiple-object tracking tasks targeting aspects of cognition including dual-tasking, covert attention, and visuomotor skill exhibited more effective eye movement strategies while driving and also showed better driving performance. Kahneman et al. (1973) found a negative association between the ability to shift attention and a driver's accident history, i.e. a lower ability is associated with increased accident frequency. Alzahabi and Becker (2013) divided their sample into light and heavy multitaskers based on their reported frequency of being simultaneously engaged in two media activities. Despite the authors found no difference regarding performance in a dual-task, heavy multitaskers were better at switching between two tasks. What is more, 2.5 % of the participants in a study by Watson and Strayer (2010) exhibited no decrease in performance when performing a difficult dual task. The authors subsequently named this group *supertaskers*. In the same study, these *supertaskers* were also found to be the top performing group in a working memory test. These findings suggest that drivers differ in their multitasking ability, namely keeping task-relevant information stored and ready, interrupting and switching tasks, simultaneously sampling and processing task-relevant information, and keeping up situation awareness. Thereby, they diverge in their potential to reach a critical out-of-the-loop state by engaging in an NDRT. As the out-of-the-loop state is assumed to yield longer response times, this should be directly visible in the drivers' take-over times. This hypothesis was tested in Article 3.

6.2 Summary of Article 3

Thirty participants took part in this study. We assessed multitasking ability by a self-developed multitasking test. This test was conducted on two separate monitors that were placed on a table at a frame-to-frame distance of 60 cm at an angle of 45° to the participant who took a seat about 40 cm in front of the monitors. On each of the two monitors, a different task had to be worked on simultaneously. The tasks were built in the Psychology Experiment Building Language (PEBL; Mueller & Piper, 2014). On the left monitor, the participants performed a reaction time task, which was a modified version of the PEBL Perceptual Vigilance Task (Karlen, Cardin, Thalmann, & Floreano, 2010; Mueller & Piper, 2014): A white fixation cross was presented for 400 ms on a black background in the center of the screen. Then, a red dot appeared at random intervals from the set [4, 5, …, 8] s. Participants had to react as fast as possible by pressing the space bar. Upon pressing the space bar, the dot disappeared and a new trial began. On the right monitor, a modified version of the PEBL Visual Search Task (Treisman, 1985) was presented. Participants had to find the letter "X" out of a random selection of 10, 20, or 30 distractor letters ("U", "D", "G", "C", "Q"). All letters were presented in white color on black background. Upon spotting the letter, the participant then had to respond with a left mouse click. Every letter previously shown was now hidden behind white circles and participants had to click on the target's location from memory. Multitasking performance as measurement outcome was defined as the sum of the reaction time of each of the two tests. The participants had to simultaneously perform both tasks for 3 min.

Körber and Bengler (2014) listed simple reaction time as another potential factor influencing take-over time. This seems reasonable since a take-over requires a fast orientation reaction and a fast response execution (see Section 2.4). Because the aim of this research was to identify the sole influence of multitasking ability on take-over time, we controlled for simple reaction time by fitting a multiple linear regression model. This allows disclosing the effect of multitasking after having accounted for simple reaction time – what is the effect of multitasking beyond the effect of simple reaction time? What does the multitasking ability tell about a participant's take-over time after his simple reaction time is known? To assess the simple reaction time, we used a modified version of the task PEBL Simple Response Time (Robinson & Tamir, 2005). The participants were presented a black letter "X" on a gray background at random inter-stimulus intervals from the set [500, 750, 1000, …, 2500] ms. Upon appearance of the letter, the participants had to press the key "X" on the keyboard as fast as possible. Each keystroke commenced a new trial with a total of 75 trials.

The participants first performed the multitasking test and the simple response task. After this, they drove for about 38 min highly automated on a highway, encountering five take-over situations. During the drive, the participants engaged in an NDRT. Potential NDRTs in automated driving will be texting with a cell phone or performing an input in an entertainment system, which are both visual-manual tasks (Horberry et al., 2006; Petermann-Stock et al., 2013). To simulate easily interruptible, visual-manual NDRT engagement, we used the Surrogate Reference Task (SuRT; detailed description in International Organization for Standardization, 2012).

We found a significant negative relationship between the performance in the multitasking test and the take-over time: The better the participants performed in the multitasking test, the lower the take-over time in the first and second take-over situation, even when we controlled for simple reaction time. The eye tracking data also reflect this finding: Participants with low multitasking test scores focused their gaze more on the NDRT and less on the road or environment. It is imaginable that participants who have difficulties performing two tasks simultaneously prioritized one task to alleviate task-induced stress. However, both relationships weakened over the course of the remaining three situations. In an exploratory analysis, we split the sample into four quartiles depending on their multitasking test performance. The mean take-over times of the first to third quartile, the 75 % best multitaskers, converged at Situation 3 and then remained on an equal level for the rest of the experiment. The participants within the second and third quartile probably either changed their task engagement strategy, increased their effort, or improved their multitasking performance by learning. However, the latter seems unlikely since the fourth quartile (the 25 % worst test performers) also experienced a reduction in take-over time. Nonetheless, an average difference of 1689 ms to the other three quartiles remained and they never achieved the same performance level of good multitaskers. It seems thus more probable that all participants adapted to the situation, but a stable difference in multitasking ability remained. Future work needs to investigate the mechanisms that determine why participants differ in their multitasking test performance and in their take-over performance development.

7 Article 4: "Introduction matters: Manipulating trust in automation and reliance in automated driving"

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction Matters: Manipulating Trust in Automation and Reliance in Automated Driving. *Applied Ergonomics*, *66*, 18–31.

7.1 The relationship between trust in automation and human-automation interaction

"There is probably no variable more important in human-automation interaction than that of trust" (Wickens et al., 2016, p. 388). First of all, trust in automation determines whether an automated system is used at all. Ghazizadeh, Lee, and Boyle (2012) state in their *Automation Acceptance Model* that trust is a pivotal determinant for an individual's acceptance of automated systems and trust is a key factor for the adoption of new technologies (Gefen, Karahanna, & Straub, 2003), the employment of automation (Lee & Moray, 1992, 1994; Parasuraman & Riley, 1997), and the intention to use autonomous vehicles (Choi & Ji, 2015). In other words, "operators tend to use automation that they trust while rejecting automation that they do not" (Körber, Baseler, & Bengler, 2018; Pop, Shrewsbury, & Durso, 2015, p. 545).

Operator and automation act as a team with assigned tasks and responsibilities while properties of both operator and automation within this team affect the human-automation interaction (Bengler, Zimmermann, Bortot, Kienle, & Damböck, 2012). Hence, "human operator trust in automation is now a major topic of interest" because it predicts not only whether automation is used but also "how automation is used" (Sheridan, 2002, p. 77). Parasuraman and Riley (1997) categorize the interaction with automation into four styles, which correspond to different levels of an operator's trust in automation: *Use* describes appropriate trust that matches the automated system's capabilities. Operator's usage and monitoring behavior lead to an enhancement in safety and improved performance of tasks that otherwise would have been performed manually (Lee, 2008). *Disuse* is present if the operator's trust lies below the automated system's capabilities. This form of inappropriate under-trust means that the operator does not accept and use automation at all or not to its full extent. Disuse is a consequence of a complex combination of the automated system's properties (e.g., frequency of failures; Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003), operator's properties (e.g., self-confidence; Lee & Moray, 1992), and situational factors (e.g., workload; Parasuraman & Riley, 1997). The third category, *Abuse*, refers to the designers of automation and describes situations in which automation is designed and deployed without taking the operator and his situation into account (Parasuraman & Riley, 1997). A mindless exchange of an operator for an automated system can cause new severe failures (Sarter & Woods, 1995).

If the operator's trust exceeds the automated system's capabilities, *Misuse* in the form of inappropriate over-reliance and over-compliance, the fourth category, may arise. Too much trust lures the operator into using automation in ways which designers did not anticipate or even try to avoid (Lee, 2008). Here, operators rely mindlessly on automation without verifying its actions and neglect monitoring the system (Hergeth et al., 2016; Muir & Moray, 1996; Rudin-Brown & Parker, 2004). Consequences are failures to detect errors (Bagheri & Jamieson, 2004; Bailey & Scerbo, 2007; Parasuraman, Molloy, & Singh, 1993) and *omission errors*, i.e. the operator does not react to a critical event if the automated system does not alert him (Meyer, 2001). Operators are then also vulnerable to automation bias (less attentiveness to contradictory information; Skitka, Mosier, & Burdick, 1999), leading to *commission errors*, i.e. an operator blindly follows an incorrect recommendation from an automated system without verification (Mosier & Skitka, 1996). Conversely, appropriately calibrated trust leads to a higher time-to-collision (Beller, Heesen, & Vollrath, 2013), a faster reaction time (Helldin, Falkman, Riveiro, & Davidsson, 2013; Seppelt & Lee, 2007), and better reaction quality (McGuirl & Sarter, 2006).

Thus, given the mentioned consequences of inappropriate trust, it is crucial to take individual differences in trust in automation into account if human-automation interaction is to be studied. The accident of the Tesla with Level 2 automation mentioned at the beginning illustrates the relevance of automation misuse caused by inappropriate trust: The driver trusted Autopilot beyond its capabilities and relied so much on it that he allegedly started to watch a movie and failed to monitor the system. It is a vivid example that trust may influence the interaction with automation in ways that are not directly connected with the design of automation (Lee & See, 2004). Consequently, the Transportation Research Board published two Research Needs Statements regarding human factors research on automated vehicles and primarily issues associated with NHTSA's Level 2 and Level 3. One of the two statements pertains to the investigation of misuse and abuse of automated vehicles (Creaser & Fitch, 2015).

7.2 Trust in automation and reliance

Based on Mayer, Davis, and Schoorman (1995), trust can be defined as "the attitude of a user to be willing to be vulnerable to the actions of an automated system based on the expectation that it will perform a particular action important to the user, irrespective of the ability to monitor or to intervene" (Körber, 2019, p. 17). It is an attitude that is closely related to reliance, which is a concrete observable behavioral outcome of trust (Körber & Bengler, 2014; Lee & See, 2004). Driving automation is a novel and intricate technology. Contrary to automation in aviation, its operators will not be professionals who have a complete comprehension of its functionality. Its operation represents a situation marked by uncertainty and vulnerability in which the user entrusts automation his well-being (Walker, Stanton, & Salmon, 2016). It is a risky, uncertain situation, where trust per definition becomes effective: If the complexity of a technology and its interaction with the situation make a complete understanding either impractical or impossible, operators tend to apply heuristics rather than analytic calculations to accommodate for their limited cognitive capacity (Gigerenzer & Selten, 2002). In such a situation, where procedures are inappropriate or cognitive resources are too limited for rational choice (Damasio, 1996), trust provides guidance for reliance and, hereby, aids operators to cope with the cognitive complexity of the system (Lee & See, 2004). Yet, trusting a system is not a binary all-or-none decision. Trusting and reliance may be rather seen as a graded process, with the degree of trust being dynamic and situational in its amount, calibration, and resolution (Hoff & Bashir, 2015; Lee & See, 2004). The object of the evaluation does not have to entirely be the system as a whole but may also be any of its particular functions. For example, trust ratings were distinct to the specific automatic controller in a study on a supervisory process control task (Lee & Moray, 1994) and an automation failure did not affect trust in the other similar but independent automatic controllers (Lee & Moray, 1992; Muir & Moray, 1996). Contrary to this, Keller and Rice (2009) presented a completely reliable aid together with an unreliable aid and report that operators tended to rate both aids the same in the sense of a global, system-wide trust rating rather than evaluating them as different, independent systems with distinct reliabilities. The definite degree of functional specificity is probably determined by an operator's experience with the system, its complexity, the information presented, and the operator's goals (Lee & See, 2004). Lee and See (2004) visualized trust as a continuum in relation to the automated system's capabilities (Figure 11).

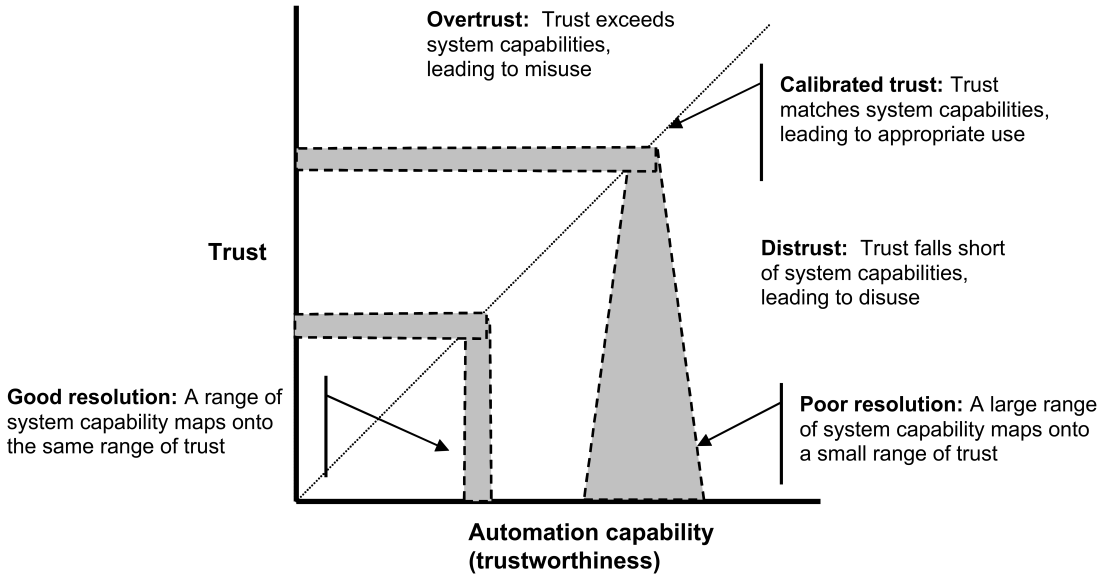

Figure 11. Appropriate trust in automation can be defined as a function of the relationship between calibration, resolution, and automation capability. Illustrated by Lee and See (2004).

7.3 Stable individual differences in trust in automation

Trust in automation is strongly dependent on the automated system in question but also exhibits a stable individual component. Although many authors have discussed what attributes constitute a trustworthy automated system, not objective characteristics but a person's subjective perception of these characteristics determines trust in automation (Lee & See, 2004; Merritt & Ilgen, 2008). Beyond this, there is evidence for a stable individual difference in how much a person is generally willing to trust a machine. In the influential model of trust by Mayer et al. (1995), trust is not only determined by the attributes of the party to be trusted (the trustee), but also depends on a person's individual trust propensity, the general willingness to trust others. Rotter (1967) defines trust entirely as an individual trait in form of a generalized expectancy that the statements or promises of another individual can be relied on – a generalization of many interpersonal experiences. The trait determines how much a person trusts a particular party prior to any knowledge of that party being available. In their meta-analysis, Colquitt, Scott, and LePine (2007) confirmed that trust propensity has a unique influence on trust beyond the trustee's trustworthiness. Merritt and Ilgen (2008) found that a user's individual perception of automation is only partially determined by the actual machine characteristics but is also influenced by the individual user's propensity to trust machines.

Mayer et al. (1995) suggest that this propensity to trust is a stable trait that varies between people, depending on their developmental experiences, personality type, and cultural backgrounds. Accordingly, in a study by Hergeth, Lorenz, Krems, and Toenert (2015), Chinese drivers reported higher automation mistrust than German drivers. Hoff and Bashir (2015) provide further evidence for cultural differences in trust in automation in their literature review. Cultural differences in trust have also been found by Schoettle and Sivak (2014a, 2014b) who surveyed about automated driving in several industrialized nations. Also, there is evidence that a form of anchoring effect exists, such as that operators lose trust upon a failure proportionally to their initial level of trust, which is in turn determined by a person's trust propensity (Lee & See, 2004; Merritt & Ilgen, 2008). For example, operators that exhibit a high expectancy that automation is trustworthy seem to be more sensitive to changes in an automated system's reliability (Pop et al., 2015). In addition to that, there is also more indirect evidence for a stable individual difference in trust. Singh et al. (1993a) argue that operators show stable individual differences in their attitude towards an automated system. The authors developed a scale to assess a person's personality in form of their potential to show complacency, which is an attentional strategy to allocate attention away from an automated task, leading to monitoring failures (Singh, Molloy, & Parasuraman, 1993b). The underlying reason for this allocation of attention may lie in their trust in the automated system. Individual differences also exist in the expectations in an automated system's reliability (Dzindolet, Pierce, Beck, & Dawe, 2002; Merritt, Unnerstall, Lee, & Huber, 2015) and Lee and Moray (1992, 1994) found strong individual differences in automation use regarding proneness to use manual control.

7.4 Attention allocation as a causal mechanism for the influence of trust in automation

Again, the question of the underlying causal mechanism arises. For example, how may it be that operators do not detect a failure or show a delayed reaction if their trust in an automated system is high? While trust in automation has various effects on the interaction with automation (Parasuraman & Riley, 1997), a possible causal mechanism of the effects (higher reaction times, more collisions) in Article 4 is described here. A Level 3 driving automation allows the driver to engage in NDRTs while the ADS is performing the driving task. By this, the driver allocates his attention away from the driving task and environment, and toward the NDRT. Previous research has already identified some problems that may arise in this scenario. *Complacency* is an attentional process that is defined as "a strategy of allocating attention away from the automated task to other concurrent tasks" (Parasuraman et al., 2008, p. 149). This strategy is probably partly a result of over-reliance on the system based on over-trust. The attentional manifestations of this strategy are an allocation of attention away from the automated task or under-sampling of the information sources that the automated system uses, based on the belief that the automated system manages the task well anyway (Wickens, Clegg, Vieane, & Sebok, 2015). For example, if an adaptive cruise control works perfectly fine in every condition, the driver may cease to pay attention whether the adaptive cruise control (ACC) keeps the right headway distance. In consequence, an operator expecting an automated system is to do its job perfectly fine will monitor it less. The consequences are monitoring failures, i.e. an operator misses or reacts extremely delayed to a malfunction, anomalous condition, or failure (Parasuraman et al., 1993; Wickens et al., 2016).

Secondly, without monitoring an operator loses awareness of the automated system's current state and its development, as well as the environment; in other words mode confusion and situation awareness, respectively (Endsley & Kiris, 1995; Manzey, 2012; Sarter & Woods, 1995). Thus, even if a failure is detected, the operator will be less likely to be able to deal with it appropriately (Wickens et al., 2016). This is particularly the case for interactions with constant highly reliable automated systems and in multitask situations (Parasuraman et al., 1993). The strategy of attention allocation may itself be partly driven by trust in the automated system and may evolve from a certain initial level of trust in automation, which is reinforced if the automated system performs at a stable high level of reliability (Parasuraman & Manzey, 2010; Wickens et al., 2016). Highly reliable systems induce trust which in turn increases an operator's reliance (Lee & Moray, 1992) and, probably, compliance with an automated system (Reichenbach, Onnasch, & Manzey, 2010), giving rise to omission and commission errors. Correspondingly, complacency most often ensues in interaction with automated systems that are perceived as highly and permanently reliable (Bahner, Hüper, & Manzey, 2008; Parasuraman & Manzey, 2010). Complacency can be reflected in two distinct types of behavioral outcomes of trust (Meyer, 2004): *Reliance* is the tendency to refrain from monitoring and acting when the automated system does not indicate the need for action. Inappropriate over-reliance promotes omission errors. *Compliance*, on the other hand, is the degree to which an operator tends to comply with an action that is instructed by the automated system. By this, over-compliance promotes commission errors (Mosier & Skitka, 1996). Trust causes attention to be shifted away from the automated task (Hergeth et al., 2016; Muir & Moray, 1996) and, consequently, a negative correlation between trust and monitoring performance has been found (Bagheri & Jamieson, 2004; Bailey & Scerbo, 2007).

The SEEV model (Wickens & McCarley, 2008) can be applied to explain the underlying mechanisms of automation complacency. In the discussion of divided attention, researchers describe attention as a limited resource that fuels conscious information processing (Neumann, 1996). However, earlier theories on attention described attention as a filter or bottleneck, which selects certain stimuli or events to be processed and filters out irrelevant stimuli (Broadbent, 1958; Kahneman, 1973; Wickens & McCarley, 2008). This filter prevents overloading the limited-capacity information processing mechanism that lies beyond the filter and processes the input in detail (e.g., its meaning; Eysenck & Keane, 2010). Yet, the exact mechanism of the filter and its location in the information processing stages are still under discussion (Sternberg et al., 2012). If a stimulus passes the filter, the information is processed under the constraint of the limited mental resources, which determines the possible activities and the number of processes that can be simultaneously carried out (Kahneman, 1973; Wickens & McCarley, 2008). The model of information processing of Wickens et al. (2016) thus only applies if the information is not filtered (Figure 12).

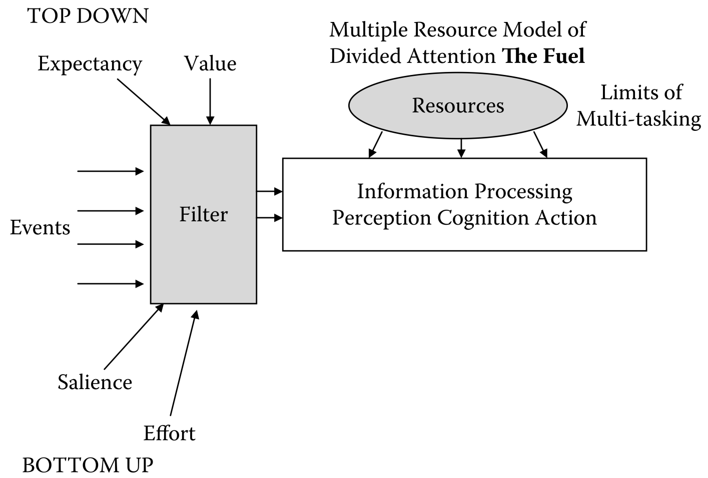

Figure 12. A simplified model of selective attention of Wickens and McCarley (2008); a filter, controlled by top-down and bottom-up factors, determines what information is processed.

The focus of attention, and, therefore, the filter, is controlled both by top-down (i.e. knowledge/expectancy-driven) and bottom-up (i.e. attention-capture) factors (Corbetta & Shulman, 2002; Wickens & McCarley, 2008). Top-down attention orienting has been described as voluntary, effortful, and controlled, whereas bottom-up orienting is an automatic process. The orientation of attention in driving is heavily top-down driven by knowledge, context, and current task goals or states such as perceived risk (Engström, 2011). For example, arriving at a T junction triggers a check of one or both directions of the road before turning because this is where relevant information concerning the turn maneuver is expected. At the same time, drivers might miss or react inadequately to unexpected events or events at unexpected locations (Shinoda, Hayhoe, & Shrivastava, 2001). According to Wickens and McCarley (2008), six factors influence visual information sampling, i.e., where our visual attentional spotlight wanders:

(1) Habit (procedural scanning)
(2) Attention capture: salience
(3) Information content: event rate or bandwidth
(4) Information content: contextual relevance
(5) Information value
(6) Effort conservation

The authors describe how these factors operate together to guide the visual attention of an operator to selectively attend a sample information at an area of interest (AoI). Based on the factors, salience, effort, expectancy, and value, the authors propose the SEEV model, which aims to predict the probability of attending to an area P(A) as follows:

$$P(A) = sS - efEF + (exEX + vV).$$

Here, *S* describes the strength of salience, which is the extent to which the AoI stands out from the background or from other AoIs (e.g., by size, color, intensity, or contrast). *EF* denotes the effort, which defines the cost of shifting attention to or away from an AoI. Salience and effort represent bottom-up influences, which can be objectively characterized by the physical environment (e.g., stimulus brightness). *EX* denotes the expectancy, i.e. the degree of how much we expect something to happen at the AoI. A high value of EX may result from a high frequency of events at the AoI, for example a neighboring lane where vehicles frequently pass by. Or, we expect an event driven by contextual cues, such as a warning or a road where we know that pedestrians frequently jaywalk. *V* represents the subjective value or the amount of information gained at this specific AoI, which may be re-described as the subjective usefulness or importance of the information. It is determined by the relevance of an AoI for a task weighted by the subjective importance of the task. Expectancy and value are top-down influences, reflecting the operator's mental model of the environment and task priorities. Wickens and McCarley (2008) provide an example for better understanding: It is crucial to detect potential collisions on the road ahead. Thereby, the AoI windshield has a high value, even if traffic is sparse (low expectancy). In contrast, the value of the AoI roadside is low, because highway advertising signs are irrelevant, albeit the expectancy (bandwidth) of the signs is high. Within this equation, the uppercase parameters describe the level of these particular parameters in a particular task while the lowercase coefficients denote the strength of these influences in directing human attention in general. Figure 13 illustrates the guidance of attention by top-down and bottom-up factors.

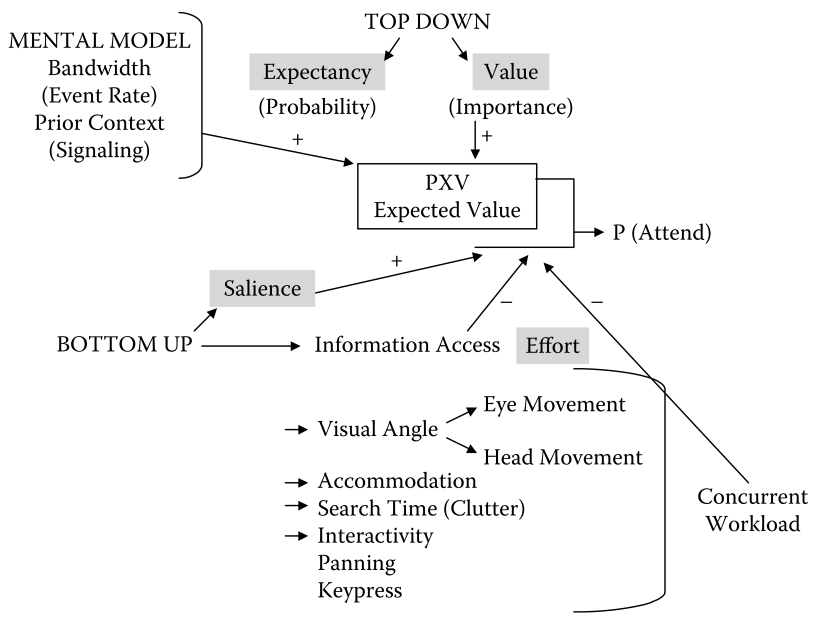

Figure 13. The SEEV model of Wickens and McCarley (2008) describes the factors that guide visual attention.

The SEEV model provides multiple reasons why even a high-value AoI, such as the automated task in complacency situations, is not attended sometimes. Firstly, the AoI has a low information bandwidth. If an automated system works very reliable, then a critical event will happen very rarely, which results in a low expectancy value for the AoI automation. Expectancy and value are multiplied in the SEEV model and the low expectancy value means for the operator that this AoI should rarely require information sampling, albeit its value for the operator is very high.

Secondly, the value could indeed also be low. Information value in the SEEV model describes an expected subjective value or the amount of information to be gained by scanning the AoI. Per definition, high trust means that the operator is willing to be vulnerable to the actions of an automated system based on the expectation that it will fulfill its tasks, irrespective of the ability to monitor or to intervene (Körber, Baseler et al., 2018). If a user completely trusts automation, he believes that the system will perfectly fulfill its tasks. Thus, monitoring the automated system becomes unimportant. Likewise, the information gained by monitoring the automated system has a low value and is of not much use for the operator. Following the SEEV formula, a low information value leads to a low probability of attending this AoI, the automated system.

Thirdly, the effort to monitor the automated system might be too high. In a dual-task situation, dual task loading competes for effort E and reduces the probability to monitor automation, especially if the automation AoI is spatially far away (Recarte & Nunes, 2000). As a result, infrequent, unexpected automation failures are difficult to detect and might be missed because the automated system is not checked anymore. Even if they are detected, operators will be less likely to respond appropriately or quickly, because they did not sample the automated system's current state, its development, or the environment (Mahr & Müller, 2011; Wickens et al., 2016). This may, at the same time, explain the *cry wolf* effect (Breznitz, 1984): After frequent false alarms, an alarm and the corresponding AoI obtain such a little SEEV value that the operator does not pay attention to the alarm anymore.

Complacency does not necessarily involve a shift in overt attention. Of course, the correlation between eye movements and attention is not perfect; unless the eyes are closed, allocation of attention to visual versus non-visual (e.g., auditory, cognitive) sources cannot be discriminated since the eyes must always fixate on something in the visual environment. However, it is still sufficiently high to be a valid indicator. Findings such as inattentional blindness (Simons & Chabris, 1999) or change blindness (Simons & Rensink, 2005) indicate that even salient stimuli in the environment may be missed albeit they are fixated. Detection or response to an object is generally improved if an individual beforehand possesses information about its features, such as location or motion (Corbetta & Shulman, 2002; Dosher & Lu, 2000; Posner, Snyder, & Davidson, 1980). For example, Posner's cueing paradigm (Posner, 1980) already showed that observers are better in detecting a stimulus in trials where it appears at an expected location (indicated by a valid cue) than control trials or trials with an invalid cue. This was the case although the participants fixated a central fixation cross for the whole time. This is an example of a *covert attention shift*, which is in contrast to *overt attention shifts*, mental focus or attention to an object without any physical movement. Indeed, there is evidence that complacency does not only consist of fixation failures (Parasuraman & Manzey, 2010). Duley, Westerman, Molloy, and Parasuraman (as cited in Parasuraman

& Manzey, 2010) obligated their participants to fixate the automated task by leaving it within foveal vision. Still, complacency in form of poorer detection performance under automation in comparison to manual control ensued.

An operator, who trusts a system and expects the automated system to successfully fulfill its tasks, probably does not expect a failure or an intervention. A lack of readiness to respond to a signal generally leads to an increase in reaction time (Poulton, 1950), while preparatory attention to a stimulus or a response generally facilitates perception as well as action (Corbetta & Shulman, 2002; Dosher & Lu, 2000). In his thorough review of driver perception-reaction times, Green (2000) comes to the conclusion that expectancy has the greatest effect. Compared to an expected event, reaction times increased about 0.5 s in common, but uncertain signals and were twice as long if the event was completely unexpected. Hence, even if operators with high trust do not divert their visual attention away from the automated task, reaction times to an event may be increased because the event is simply not expected. Expectation guides attention and leads to overlooking subjectively less probable events (Koustanaï et al., 2008). In such *looked-but-failed-to-see* accidents (Herslund & Jørgensen, 2003), drivers fail to detect an obvious hazard although they were looking in its direction. It follows that neither high trust nor low trust but appropriately calibrated trust is important. To raise the expectation of a possible upcoming event, the uncertainty of an ADS may be visualized. Indeed, studies have shown that the presentation of uncertainty improved situation awareness and led to an increase in time-to-collision in the event of an automation failure (Beller et al., 2013). Drivers in the study of Helldin et al. (2013), who were provided with an uncertainty representation, regained control of the vehicle faster when needed.

7.5 Summary of Article 4

Trust in automation does not only predict whether an automated system is used but also how it is used (Parasuraman & Riley, 1997). Inappropriate over-trust and the resulting over-reliance has been associated with monitoring failures, longer reaction times, and poorer reaction quality (Bagheri & Jamieson, 2004; Bailey & Scerbo, 2007; Beller et al., 2013; Helldin et al., 2013; McGuirl & Sarter, 2006; de Waard, van der Hulst, Hoedemaeker, & Brookhuis, 1999). An operator should be aware of the capabilities and should rely on it adequately when it is close to the limits of its capability (Carlson, Drury, Desai, Kwak, & Yanco, 2014). Otherwise unexpected situations may occur in which the reaction to unexpected events is known to be delayed or poorer (Green, 2000; Wickens et al., 2016). In conditionally automated driving, the take-over of vehicle control can then pose a critical situation if the ADS is operated in an unfamiliar, unexpected, or unstructured environment, situation, or condition (Shinar, Tractinsky, & Compton, 2005; Wagner & Koopman, 2015). For example, Payre, Cestac, and Delhomme (2016) found a positive correlation between take-over time and trust if training was insufficient. Consistent across different levels of automation, an inappropriate level of trust leads to longer reaction times or inferior reaction quality (Abe, Itoh, & Tanaka, 2002; McGuirl & Sarter, 2006; Parasuraman & Riley, 1997). Some of the mentioned findings have been studied in non-driving-related automation paradigms, such as supervisory control tasks. The goal of this study was to assess the relationship between trust, reliance, and take-over performance in the context of conditionally automated driving in greater detail. More precisely, we investigated if trust is associated with monitoring of the ADS, reliance on the ADS, take-over time, and take-over quality.

In contrast to aviation, the drivers of automated vehicles will not be professional experts but laypeople (Casner, Hutchins, & Norman, 2016). Automated driving, including the engagement in NDRTs, represents a novel situation for future drivers. As already mentioned, trust is particularly important here because it determines reliance in complex, novel, or unanticipated situations that are not or cannot be completely understood (Lee & See, 2004; Meyer, 2001). In such a situation, instructions and training with automation form expectations, prior knowledge, and understanding (Hoff & Bashir, 2015). How reliable an operator expects an automated system to be depends very much on how an automated system is presented (Barg-Walkow & Rogers, 2016; Mayer, Rogers, & Fisk, 2009). This expectation, in turn, determines how an automated system's reliability is perceived (Madhavan & Wiegmann, 2005; Pop et al., 2015). Explicit statements about system reliability in the introduction set the initial reliance, compliance, and tendency towards over-reliance (Madhavan & Wiegmann, 2005; Mayer, Sanchez, Fisk, & Rogers, 2006). Accordingly, Beggiato and Krems (2013), for instance, reported a negative relationship between the preliminary presented number of potentially critical situations and initial trust in an adaptive cruise control system (ACC).

In this study, we aimed to transfer the mentioned findings to conditionally automated driving and investigated whether trust-promoting and trust-lowering introductory information influences reported trust, reliance behavior, and take-over performance. We expected that an ADS that raises positive expectations regarding its performance (Schaefer, Chen, Szalma, & Hancock, 2016), signals high reliability (Mayer et al., 2006), and is experienced as reliable (Sauer, Chavaillaz, & Wastell,

2015) should promote initial trust in automation. Forty participants were equally assigned to two groups (*Trust Lowered* or *Trust Promoted*) that received either trust-lowering or trust-promoting information prior to an experimental drive. Trust in automation was measured by the TiA questionnaire, which was developed in the course of this study and is discussed in Section A4 in the appendix. The participants then encountered three situations during a 17 min highway drive in a conditionally automated vehicle (SAE Level 3). Situation 1 and Situation 3 were non-critical situations that were solved by the ADS, but a take-over was a reasonable action if one did not trust driving automation. Situation 2 represented a critical situation where a take-over by the driver was necessary to avoid a collision. An NDRT, the SuRT, was presented between the situations to make participants' allocation of visual attention observable. Results showed that participants reporting a higher trust level spent less time monitoring the road or instrument cluster and more time looking at the NDRT. The manipulation of introductory information resulted in medium differences in reported trust. Accordingly, participants of the *Trust promoted* group monitored the road or instrument cluster less and looked more at the NDRT. The odds of participants of the *Trust promoted* group to intervene, i.e. to overrule the ADS, in the non-critical situations were 3.65 times (Situation 1) to 5 times (Situation 3) higher. In the critical take-over situation, the *Trust promoted* group's mean take-over time was 1154 ms higher and the mean minimum time-to-collision was 933 ms lower. Six participants from the *Trust promoted* group compared to zero participants of the *Trust lowered* group collided with the obstacle.

The results highlight that the individual trust level influences not only how much drivers monitor the road and the environment while being engaged in an NDRT, but also the reliance on an ADS, and predicts if a critical take-over situation can be successfully solved. Introductory information influences this trust level. However, it has to be taken into account that the study was conducted in a driving simulator because of the potential risk of collisions and injuries. It is imaginable that the participants may have behaved in a riskier manner in this setting because of the lack of real negative consequences. The demonstrated impact the information material has on the participants' trust in automation implicates that car manufacturers should carefully consider how they present and introduce an ADS to their customers.

8 Article 5: "Vigilance, boredom proneness and detection time of a malfunction in partially automated driving"

Körber, M., Schneider, W., & Zimmermann, M. (2015). Vigilance, boredom proneness and detection time of a malfunction in partially automated driving. In *International Conference on Collaboration Technologies and Systems (CTS)* (pp. 70–76). IEEE.

8.1 Partially automated driving as a vigilance task

In Level 2 automation, i.e. partially automated driving, the longitudinal and lateral control of the dynamic driving task is transferred to a driving automation system, but the driver has to constantly monitor the environment and the system. In case of an event, the driver has to be ready to immediately execute an appropriate response (SAE International, 2016). Hence, the driver's role shifts from an active controller to a passive monitor. Since the driver is being relieved from controlling the vehicle, it seems to be a logical consequence that now enough attentional resources are available to monitor the driving automation system. However, this shift in vehicle control still necessitates further investigation since past research has shown that humans are low performers in monitoring tasks and in monotonous situations with low task demands (Molloy & Parasuraman, 1996; Pattyn, Neyt, Henderickx, & Soetens, 2008; Saxby et al., 2008). In the case of a reliable automation, malfunctions are rare events (Parasuraman et al., 1993). To detect a malfunction, the operator has to sustain his attention and monitor the automated system for a long time without any critical event occurring. Sustained attention, also called vigilance, is a state or degree of readiness to detect and to react to rarely occurring stimuli that appear at random intervals for an extended time (Körber, Schneider, & Zimmermann, 2015). Although the task might seem easy, humans generally perform poorly in vigilance tasks and detection performance declines in the course of these tasks (Parasuraman, 2009). For example, Molloy and Parasuraman (1996) found that more participants detected an automation malfunction in the first 10 min than in the last 10 min of a 30 min session. Parasuraman and Riley (1997) collected further findings on low monitoring performance when automation reliability is high, which creates a monotonous monitoring situation. A highway drive can be monotonous and detrimental to the driver's arousal level, conditions that are also present in vigilance tasks. Such a monotonous situation has been already shown to reduce vigilance in manual driving (Larue, Rakotonirainy, & Pettitt, 2011; Schmidt et al., 2009; Thiffault & Bergeron, 2003).

Partially automated driving may be seen as a vigilance task because the operator has to ignore non-critical events for a long time and has to react to comparably rare critical events. Studies on vigilance tasks and automation monitoring tasks have shown that individuals differ in their ability to sustain attention and to monitor an automated system (Finomore, Matthews, Shaw, & Warm, 2009; Körber & Bengler, 2014; Matthews, Warm, Shaw, & Finomore, 2010; Prinzel, DeVries, Freeman, & Mikulka, 2001; Shaw et al., 2010; Singh et al., 1993b). It follows that not only the situation but also the individual operator is relevant for the resulting monitoring performance. The decrement of vigilance during a partially automated highway drive has already been investigated by Körber, Cingel et al. (2015), which is discussed in detail in the complementary article provided in the appendix, Section A2. Building on this, we investigated if a person's ability to stay vigilant is directly related to an individual's detection performance of a malfunction of a driving automation system.

Furthermore, since relevant events for the driver occur with low frequency, monitoring tasks are generally seen as boring (Scerbo, 1998; Straussberger, 2006). Following the mindlessness theory (Smallwood & Schooler, 2006), a situation where low stimulation leads to a state in which the mind begins to wander, task-irrelevant thoughts and inner monologues occur. The driver is lost in thoughts and exhibits an increased variability of reaction times (Seli, Cheyne, & Smilek, 2013) and takes longer to respond to sudden events (Yanko & Spalek, 2013). However, how fast a situation becomes boring depends on the individual operator: Farmer and Sundberg (1986) showed that there are also individual differences in how fast individuals get bored by a situation. They introduced the construct *boredom proneness*, which describes the extent of stimulation needed to keep a person from becoming bored. Therefore, we expect that a driver will get bored when driving automation is activated, and engages in mind wandering and task-irrelevant thoughts, which may increase reaction times. This hypothesis was also investigated in this study.

8.2 Summary of Article 5

We conducted a driving simulator experiment with 23 participants who drove for 24 min on a three-lane highway with partial automation activated. To assess individual vigilance performance, we developed a vigilance task in E-Prime 2.0 based on the short vigilance task by Temple et al. (2000) and similar tasks (Caggiano & Parasuraman, 2004; Helton & Russell, 2013). A fixation cross appeared for 200 ms accompanied by two zeros left and right of it. Both zeros had the same distance of 25 mm to the fixation cross in noise trials. In signal trials, one of the two zeros horizontally deviated 5 mm from its position. Participants should react as fast as possible if they spotted a signal trial. No reaction was required in noise trials. A total of 600 trials were presented which resulted in a total test time of about 12 min. Boredom proneness was measured by the Boredom Proneness Scale (BPS; Farmer & Sundberg, 1986).

The dependent variable was the reaction time to an automation malfunction. This malfunction was implemented as an extensive right curve. At the beginning of this curve, the vehicle's lateral control malfunctioned without noticing the driver. Given these circumstances, the car went straight into the long right curve and eventually left the middle lane, in which the participants were supposed to drive. The dependent variable was the time until they detected this suddenly occurring malfunction of lateral control.

The vigilance test developed for this study showed similar results as comparable tests. No significant relationship between the predictors and reaction time was found, although boredom proneness seems to be a promising predictor with a standardized regression coefficient of Beta = 0.25. Reasons for the non-significant effects could be the chosen operationalization of the dependent variable. The main determining factors for reaction time were probably the location of visual attention at this moment and the individual threshold to classify the deviance from the lane center as a malfunction. These two factors might have outweighed the influence of personality traits. Further explanations are the study's small sample size, which implies a low statistical power, or the novelty of the driving simulator for the participants, which may have eliminated the monotony and boredom of the situation.

9 Where to go from here? General discussion, limitations, and future work

Following the presentation of the articles of this thesis, this chapter provides a more general discussion of the findings and the methods used, as well as potential objectives for further research. Based on the results, practical implications in form of recommendations on courses of action are given.

9.1 Trust in automation: More than a feeling

Trust is a latent, non-observable construct, yet as reported in Article 4, it has severe observable consequences: Six participants (30 %) of the *Trust Promoted* group compared to zero participants of the *Trust Lowered* group collided with an obstacle in a take-over situation. A participant's level of trust furthermore predicted the interaction with the ADS while being engaged in an NDRT. As expected, participants who reported a higher level of trust prior to the experimental drive then spent less time looking at the road and the instrument cluster and more time looking at the NDRT. At the same time, participants with lowered trust showed traces of automation disuse when they unnecessarily overruled the automated system. Thus, trust in automation is a significant determinant of a driver's interaction with an ADS: It influences how much drivers monitor the environment while performing an NDRT, reliance on the ADS, and if a critical take-over situation is solved successfully. Beyond that, the study showed that the trust level and thereby the interaction with an ADS can be influenced by the provision of introductory information. The results empirically confirm the theoretical predictions and transfer the findings on the role of trust in the interaction with automation to the field of automated driving. Yet, the generalizability of the results is certainly not without limits. Risk is absent in a driving simulator and, therefore, the experimental situation might have elicited riskier behavior than during a naturalistic drive. We conducted the study in a high-fidelity driving simulator and chose a realistic scenario regarding the deployment of driving automation. We still believe that the participants' immersion was sufficient to ensure that their behavior in outlines reflects naturalistic behavior. While the sample size per group was rather small, we found the same relationship between trust and reliance consistently over the three situations, which indicates a stable effect. Further evidential value comes from the fact that the study embodied a confirmatory research design, i.e. the findings were predicted from the theory on trust in automation (for details see Section 7.2) and are consistent with related work (Hergeth et al., 2016; Manzey, 2012; Parasuraman & Riley, 1997).

Because the presented introductory information material systematically influenced participants' trust in automation, the resulting recommendation is that car manufacturers should carefully consider how they present the ADS to customers. Trust generally calibrates according to experiences made with the automated system. It usually increases up to a stable high level if the system performs reliably and as expected (Beggiato, Pereira, Petzoldt, & Krems, 2015; Hergeth et al., 2016). If operators are unaware of a system limit, its first occurrence triggers a calibration process. But this calibration comes too late for the first occurrence of a malfunction and over-reliance may have already been present. Neglecting the limitations of an ADS at its presentation and at the same time focusing on its capabilities may lead to miscalibrated trust and over-reliance, at least with little further experience with the system. On the other hand, automation failures generally result in a drop of trust but its impact depends rather on its predictability than on its magnitude (Lewandowsky, Mundy, & Tan, 2000). A failure erodes trust if limitations are omitted in the system introduction given beforehand (Beggiato & Krems, 2013). But when the failure is predictable or when its cause is comprehensible, no drop in trust occurs (Dzindolet et al., 2003). Therefore, the correct introduction of an ADS is crucial not only to avoid over-reliance but also to ensure acceptance. The Tesla accident shows that the provision of information material alone may not be sufficient: Over-reliance occurred albeit all necessary information on Tesla's Autopilot was provided by the company (Office of Defects Investigation, 2017). The majority of future users of automated vehicles will not be professionals but laypeople without detailed knowledge of driving automation (Walker et al., 2016). At the beginning, they will construct a mental model of the system's functioning based on either the information provided with the system or their interactions with it (Naujoks & Totzke, 2014). Without the provision of correct and comprehensible feedback, customers might construct an incorrect mental model (Christoffersen & Woods, 2002). The recommendation derived from this study is not only to provide information material and training but also to pay close attention to its content and presentation. The training or the information material should describe how particular situations interact with the characteristics of the ADS to influence its capability to solve the situation (Lee & See, 2004). Thus, customer-tailored training and information material as an introduction to automated driving could appropriately calibrate trust and could mitigate the negative effects of an automation failure (Wickens & Xu, 2002). Further thoughts on the calibration of trust by design of the human-machine interface (HMI) or a gamification approach (Robson, Plangger, Kietzmann, McCarthy, & Pitt, 2015) can be found in the corresponding article by Körber, Baseler et al. (2018) and in the complementary article in Section A1 of the appendix.

In the same manner as car manufacturers, experimenters need to consider how they present the ADS to their participants. An overly optimistic presentation without acknowledging the limitations or a too narrow focus on them may lead to miscalibrated trust. If the majority of participants collides with an obstacle because of over-reliance, the influence of an HMI is difficult to assess. Furthermore, researchers may include a measurement of trust in their experimental design to explain unusual behavior, possibly combined with an interview. A questionnaire is an attractive option for measurement trust in automation. For this reason, a questionnaire was developed in the course of the study. Its development process and the evaluation of its psychometric quality are discussed in Section A4 in the appendix. If a strong influence of trust is assumed, researchers can include reported trust in automation as a covariate in their regression model to reduce error variance in take-over times.

In the study, trust in automation was measured before the experimental drive. This highlights the predictive validity and influence of trust on the subsequent interaction with driving automation. In other words, miscalibrated trust is not only a post hoc diagnosis in case of an accident, trust measures can also to a certain extent predict the subsequent interaction with automation. However, trust is not the only factor determining the interaction. Even though trust is high, over-reliance does not follow with certainty since, as mentioned previously, attitudes and real behavior are in a complex, non-deterministic relationship. As a consequence, the interaction with automation cannot be predicted with absolute certainty, even with moment-to-moment measurements of trust (Drnec, Marathe, Lukos, & Metcalfe, 2016). Affinity to technology and trust in automation are in a close relationship and were consequently found to highly correlate in this study. Although their exact inter-relationships remain to be elucidated, both, together with prior or expert knowledge on automation, seem to have a strong influence on the interaction with automation, especially in the case of the first real interaction. Since they are at least partly stable individual characteristics, it is essential for researchers to carefully consider how they constitute their study sample according to their research goal and the population they want to infer to.

The study investigated the relationship between trust in automation and reliance behavior in form of gaze behavior and voluntary interventions. The participants were instructed to engage in the NDRT during the drive. Future research can build on these results and could investigate the relationship between trust in automation and voluntary engagement in NDRTs. Participants with a higher trust level could be more comfortable with voluntary engagement in an NDRT and might begin to engage with it in a shorter time. Generally, many predictions can be derived from the hypothesized relationships between trust and the interaction with automation. However, only few empirical studies have been performed yet that explicitly test these predictions. Replications and effect size estimation are required to stabilize the theory's nomological network.

However, the probably most important question to answer is the development of trust in long-term use of automated vehicles and the possible development of complacency. In longitudinal studies, researchers now need to investigate how trust in automation develops in day-to-day use, especially over a longer period of time. The study's results and Brown's Tesla accident mentioned in the beginning emphasize the urgency of this research question. Section 9.7 provides a broader discussion of this issue as well as recommendations on suitable methods.

9.2 What is the influence of age in automated driving?

Aging influences information processing in the model of Wickens et al. (2016) at every stage and in each component (Anstey et al., 2005; Bryan & Luszcz, 2000; Der & Deary, 2006; Miller et al., 2016; Salthouse, 2009). However, we found no significant difference in the take-over performance between younger (≤ 28 years) and elderly (≥ 60 years) drivers. It seems that an age-related cognitive decline measured in laboratory tasks does not invariably translate into a functional decline of everyday performance (Green, 2000; Park & Gutchess, 2012; Wickens & McCarley, 2008). Laboratory tests are set in a very controlled environment that provides the sensitivity to detect differences that might be washed-out in naturalistic settings by confounding factors such as strategies, experience, and expectancy. Although everyday tasks and situations may at first appear to be cognitively demanding on first sight, the majority of driving tasks is immensely well-trained and, hence, handled by automatic rather than conscious effortful processing (Schneider & Shiffrin, 1977). Moreover, to compensate the effects of aging, elderly people develop task-related strategies such as prioritizing one task in a multi-task situation or taking notes in memory-heavy tasks

(Andrews & Westerman, 2012; Wickens & McCarley, 2008). By this tactical compensation, they allow themselves more time to react to events and to make driving-related decisions (Andrews & Westerman, 2012). For example, elderly drivers drive more slowly (Hakamies-Blomqvist, 1998) and choose bigger gaps at junctions (Middleton, Westwood, Robson, & Kok, 2005). Elderly drivers also avoid driving conditions and traffic situations that do not suit their abilities (Baldock, Mathias, McLean, & Berndt, 2006; Molnar & Eby, 2008). It is imaginable that any potential difference in cognitive functioning could have been neutralized by compensational strategies and experience.

The results also have to be interpreted in light of the potential confound trust in automation. Previous studies gathered weakly informative, mixed support for a relationship between age and trust in automation (Gold, Körber, Hohenberger, Lechner, & Bengler, 2015; Ho, Wheatley, & Scialfa, 2005; Sanchez, Fisk, & Rogers, 2004). Surveys on automated driving have reported a slight tendency of elderly towards having a more negative attitude towards automated driving (Abraham et al., 2018; Payre, Cestac, & Delhomme, 2014; Schoettle & Sivak, 2016). Their attitude towards technology, in general, seems to be more negative (Ellis & Allaire, 1999), however, the relationships seem to very complex with multiple moderating and mediating variables present (Czaja et al., 2006). Hoff and Bashir (2015) suggest that people of different ages could employ different strategies in their analysis of the trustworthiness of an automated system, while the specific effect of age may vary according to the context. Donmez, Boyle, Lee, and McGehee (2006) also state the effect of age may strongly depend on the type of technology assessed. Furthermore, individual age is confounded with the expertise with a technology, which in turn is related to trust (Rudin-Brown & Parker, 2004). If older drivers, compared to younger drivers, differed in their trust in automation and their attitude towards driving automation systems, they might have handled the take-over situation more seriously and with greater attention. Moreover, it is imaginable that, given their more skeptical view on automated driving, the elderly participants could also have taken the driving simulator experiment more seriously than the younger drivers. It is imaginable that this potential difference in trust in automation and attitude towards technology could have neutralized any existing difference in take-over performance.

Despite this, the results seem to suggest that the driver's age has no relevance in automated driving. This finding is in line with other studies that used different scenarios (Naujoks, Purucker, Neukum, Wolter, & Steiger, 2015; Petermann-Stock et al., 2013) and has since been conceptually replicated in a more recent study (Clark & Feng, 2017). One can therefore quite certainly state that age does not seem to influence take-over time in driving simulator studies in similar, yet slightly different conditions. The question is to what extent the results can be generalized. Is the conclusion "age has generally no influence" valid? This depends on the external validity of a study. The external validity is "the extent to which the results of research or testing can be generalized beyond the sample that generated them" (VandenBos, 2015, p. 402). High external validity means that a relationship that has been demonstrated in one specific research setting can be obtained in other settings, at different times, with a different sample, or different research procedures (Brewer & Crano, 2014). Many textbooks underline the limits of the generalizability of results regarding the sample of participants and the inference from a sample to a population (Hammond, 1986). For example, if a study has been conducted solely with male participants, the results may not generalize to female participants (VandenBos, 2015). It is debatable if the results of this study generalize to the whole population of elderly drivers since the cognitive decline is highly variable in its degree and depends on personal lifestyle, experiences, and genetics (Deary et al., 2009; Hultsch et al., 2002; Morse, 1993). High inter-individual variability in these factors consequently leads to a variability in the cognitive functioning as well. Not aging per se seems to cause an increased crash risk but a decline in particular abilities (Hakamies-Blomqvist, 1998). Same-aged participants may thus differ strongly in their performance. Given this high variability, it has to be taken into account that the participants in this study were volunteers who probably considered themselves fit enough to perform a novel task in a driving simulator. The sample may therefore not represent a random sample of the whole population of elderly drivers but a random sample of the population of elderly drivers who are affine to technology and who think of themselves as cognitively fit. This sample bias may skew the difference between younger and elderly drivers. Taken together, these results allow the conclusion that no significant difference among cognitively fit elderly and younger drivers exists, but they do not guarantee this statement to be valid for the whole population of elderly drivers. At the same time, a non-significant mean difference does not imply that a small proportion of elderly with high cognitive decline can still safely interact with a driving automation system – in the end, it is merely an average difference.

Another limitation stems out of the fact that the non-significant difference was observed in a driving simulator under highly controlled conditions. This limits the interpretation of the results in two ways. Firstly, previous research has confirmed relative validity for certain tasks in high fidelity driving simulators such as the driving simulator used in this study. That means that the drives in a simulator and on-road produce values similar in magnitude and with the same sign (Blana, 1996; Mullen, Charlton, Devlin, & Bédard, 2011). But this validity is specific to each task and condition. Brunswik (1956) stresses that not only the sample but also the conditions of the experiment must be representative of the context that the researcher wants to make inferences about. He sees ecological validity, the degree to which experimental conditions are representative of conditions in the real world, as a crucial fundament for external validity. Designing situations that are convenient to analyze or produce a large effect but are atypical for the individual, here the driver, lowers the ecological validity. Low ecological validity may create clear findings that are nevertheless not observable outside of the experiment. The focus on high internal validity may sometimes come at the expense of external validity, the generalizability of a causal relationship beyond the conditions under which it was studied or observed. It is crucial in the design of an experiment to always consider its relevance regarding conditions beyond the laboratory setting. The experimenter should know and understand the corresponding real-world context and tasks which should lead the design of the experiment (Wickens et al., 2016). Otherwise, researchers end up "with results confined to a self-created ivory tower ecology" (Brunswik, 1956, p. 110). A driving simulator provides limited ecological validity because, contrary to a naturalistic drive, risk and real consequences are absent, which may result in riskier behavior. For the majority of participants, driving with an automated vehicle constitutes a novel situation contrary to the routine that is inherent in the vast majority of drives. Furthermore, participants know that they are monitored and may be extra-motivated or act socially desired (Kircher, 2007). We implemented realistic take-over situations and an NDRT that emulates a phone call. The driving simulator was a high-fidelity mockup and the experimental track was a realistic highway drive, which is the most likely scenario for the first deployment of a driving automation system. At the beginning of the experiment, participants were instructed to obey traffic rules and to ensure safe driving at any time. We believe that the participants' behavior and perception did not substantially differ from a naturalistic drive and that meaningful results were obtained. Therefore, the experimental design itself exhibited sufficient ecological validity. However, the experiment still comprised a very controlled situation which differed from a naturalistic drive in some aspects. For example, the drivers were prompted to engage in the NDRT throughout the whole experimental drive. This has to be taken into account regarding the generalizability of the results.

Secondly, an experiment itself can also be thought of as a sample, namely, a sample of behavior. Again, the characteristics of the sample determine the population to which it can be inferred (Hogarth, 2005). The interpretation and generalizability of findings also depend on the stimuli and the conditions of the experiment (Fiedler, 2011), which are also just a sample of the population of relevant conditions or possible stimuli. Just like the sample of participants determines the population to which the results can be inferred, the sample of conditions also limits the conditions to which the results generalize. The stimuli should be representative of a defined population of stimuli, for example in terms of their number, distribution etc. To obtain ecological validity, a researcher must use stimuli that are representative of the natural stimuli to which the experimenter wishes to generalize (Dhami, Hertwig, & Hoffrage, 2004). The conditions of the take-over situation in this experiment were far from optimal: Different traffic densities and an NDRT were implemented. Both age groups did not differ in their take-over time and suffered from higher traffic densities to the same extent. Still, to generalize this finding universally to all possible conditions would be too big of a leap. It is conceivable that younger drivers' capabilities were not exhausted to the full extent in the take-over situation while elderly participants were already operating at their limits. With increasing experience, driving becomes a more and more automated task, which is known to be executed with minimal cognitive effort (Schneider & Shiffrin, 1977). Experience compensates for differences in the driving ability in routine situations, but differences appear in novel situations (Baltes, Staudinger, & Lindenberger, 1999). Elderly drivers are overrepresented in crashes occurring from complex situations, such as junctions, where self-paced task performance is not possible (Andrews & Westerman, 2012). If the conditions get worse – a shorter take-over time or a more complex response may be required – differences between the groups may appear. For example, in the study of Bélanger, Gagnon, and Stinchcombe (2015), elderly drivers showed a higher crash risk than younger drivers when a critical event required multiple synchronized reactions.

In conclusion, this study, in combination with the other studies (Clark & Feng, 2017; Naujoks et al., 2015; Petermann-Stock et al., 2013), provides initial evidence that age does not have a significant influence on the take-over performance in a driving simulator. However, more evidence is needed to provide a clearer and more general conclusion. In addition, a naturalistic drive could confirm the ecological validity of the findings. The challenge now is to identify the reason for the contradiction between findings on cognitive decline and the study's finding of no significant difference between younger and elderly drivers in take-over time by eliminating competing explanations (Platt, 1964). Future studies should recruit a more diverse sample of elderly participants to rule out the effect of the mentioned sample bias. An implementation of more difficult or more elaborate scenarios, where a more complex decision or reaction is required, could test the hypothesized compensation strategies. Differences in trust or in attitudes can be measured by questionnaires and implemented as covariates.

9.3 Dealing with non-significant results: Equivalence testing to accept the null hypothesis

The data in Article 2 showed no significant difference between the two age groups; we did not reject the null hypothesis of no difference. What conclusion can be drawn? The *p*-value is only an indirect indicator of evidence against the null hypothesis and states how compatible the data are with it. It would be a fallacy to conclude that the groups are equal, i.e. no difference exists, based on an observed high (non-significant) *p*-value alone. The typical NHST procedure states a null hypothesis that is supposed to be rejected by the data. The reasoning then relies on modus tollens, denying the consequent. That means that if the null hypothesis is rejected, the complementary alternative hypothesis is accepted. If the data are not surprising enough, assuming the null hypothesis is true, we cannot accept the null hypothesis but only fail to reject. Because of this, a non-significant result cannot be taken as evidence for the null hypothesis. This is true because of several reasons (Greenland et al., 2016): Firstly, accepting the null hypothesis would ignore the principle "absence of evidence is not evidence for absence" (Oliver & Billingham, 1971, p. 5). The *p*-value can only state evidence against the null hypothesis but not for it. Insufficient evidence for a rejection does not imply sufficient evidence to accept the null hypothesis (Blackwelder, 1982). Secondly, the data may not be informative either for the null hypothesis or for the alternative hypothesis. Moreover, even if the null hypothesis is wrong, a *p*-value may be large because of a large random error term or another violated assumption. Thirdly, a large *p*-value merely states that the data are compatible with the null hypothesis, but many other alternative hypotheses may be equally or more compatible with the data (Wasserstein & Lazar, 2016). Fourthly, the high *p*-value only states that, given a certain sample size, the difference may not be large, but it is no evidence for a difference of zero (Greenland et al., 2016).

After deciding to not reject, a Bayes factor (BF) could be calculated to quantify how sensitive the data distinguish the null hypothesis from the alternative hypothesis (Dienes, 2014; Rouder, Speckman, Sun, Morey, & Iverson, 2009). A BF is the ratio of the probability of the data given a null model to their probability given an alternative model and, by that, quantifies whether the data are more compatible with the null model or the alternative model (Körber, Prasch et al., 2018; Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2017). The null model, in this case, would be no difference between the groups. A BF, contrary to a *p*-value, is able to obtain evidence for the null hypothesis and, thereby, can distinguish between uninformative results and results supporting the null hypothesis (Dienes, 2014). However, an alternative model has to be specified and implemented as prior distribution for the expected effect. Thus, the BF would tell us if the data are more compatible with a difference of 0 ms or with, for example, a normal distribution on an effect of $d = 0.10$. Hence, the results strongly depend on the chosen alternative hypothesis. If the aim is concluding that there is practically no difference, an equivalence test approach may be me more informative (Lakens, 2017). An equivalence test can be performed either by performing two one-sided tests (TOST procedure) or by interval estimation – both procedures lead to the same conclusion. In this procedure, the null hypothesis of an effect size deemed relevant is supposed to be rejected. An equivalence region $[-\delta, \delta]$ has to be proposed. It may be based on a consideration regarding effect sizes, for example a small effect of $d = 0.10$. Alternatively, theoretical considerations may guide the choice. For example, given a maximum test result of 100 score points, a difference of ± 3 score points in a test on mathematics ability may be deemed as practical irrelevant, i.e. indistinguishable and essentially the same ability.

An equivalence region for the take-over time may be set to ± 300 ms, the approximate mean simple reaction time to a visual stimulus (Deary & Der, 2005). Thus, equivalence would be rejected when the groups differ more than the time it takes to react. Another equivalence boundary could be set from an accident related perspective. At a speed of 100 km/h on a highway, a vehicle would take 180 ms to drive 5 m. Thus, equivalence would be established if the distance between two entities of each group at the brake reaction point would be on average less than approximately 5 m. Interval estimation provides an illustrative way to investigate if the two age groups are equivalent regarding their take-over time. The procedure quantifies the uncertainty of a parameter estimation and thereby answers how sure we can be about the estimated mean difference. Interval estimation can be realized for this purpose in form of a frequentist confidence interval or a Bayesian credible interval. The possible results of an equivalence test are illustrated in Figure 14. If the interval boundaries of the difference are within the equivalence boundaries, equivalence is established (Case A and C). If not, we can conclude that we cannot tell if they are either different or equivalent with the given data – the data are inconclusive (Case D). More, informative data has to be gathered in this case. The confidence interval is the 90 % confidence interval based on the *t* distribution (Walker & Nowacki, 2011). The priors for the credible interval are similar to

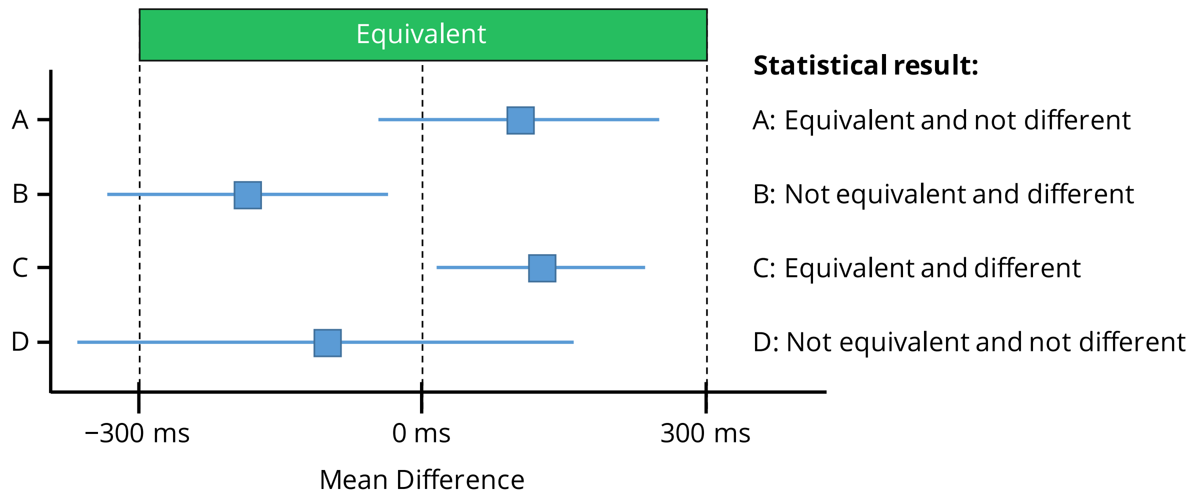

Figure 14. Possible results of a statistical test on equivalence and difference; adapted from Lakens (2017).

Körber, Radlmayr, and Bengler (2016) and again scripts by Kruschke (2015) were used. For the estimation of the difference in take-over times y_{diff}, we chose generic, relatively uninformed, and robust prior distributions that are only minimally informed by the scale of the data. Each group mean μ_j was estimated with a normal prior distribution that was adjusted to the sample data. The group standard deviation σ_j was estimated with a broad flat uninformed prior with boundaries adjusted to the sample standard deviation. Take-over times of both groups y_{ij} were assumed to come from a t distribution with a mean μ_j, standard deviation σ_j, and normality parameter ν:

```
μj  ~ normal(mean(y), 1/(SD(y)·100)²)
σj  ~ uniform(SD(y)/1000, SD(y)·1000)
ν   ~ exponential(1/30)
yij ~ t(μj, 1/σj², ν)
```

The estimated mean difference between the two age groups and their corresponding 90 % confidence intervals/credible intervals are listed in Table 2 and illustrated in Figure 15. Given an equivalence region of ± 300 ms or ± 180 ms, equivalence cannot be established in both cases. It follows that the mean difference in take-over time between both age groups is neither significant nor are the groups equivalent. That means the data of this study are inconclusive regarding the influence of age on take-over time. They do not provide sufficient evidence for a difference among younger and elderly drivers, but at the same time they do not allow the statement that the found difference is negligible. Taken together, what is the verdict on the influence of age on take-over

Table 2
Resulting confidence intervals and credible intervals for the mean difference in take-over times at different traffic densities

	Mean Difference [CI]	Mean Difference [CrI]
Traffic Density 0	184 [−222, 589]	236 [−200, 640]
Traffic Density 10	177 [−334, 688]	173 [−366, 698]
Traffic Density 20	153 [−324, 630]	132 [−362, 631]

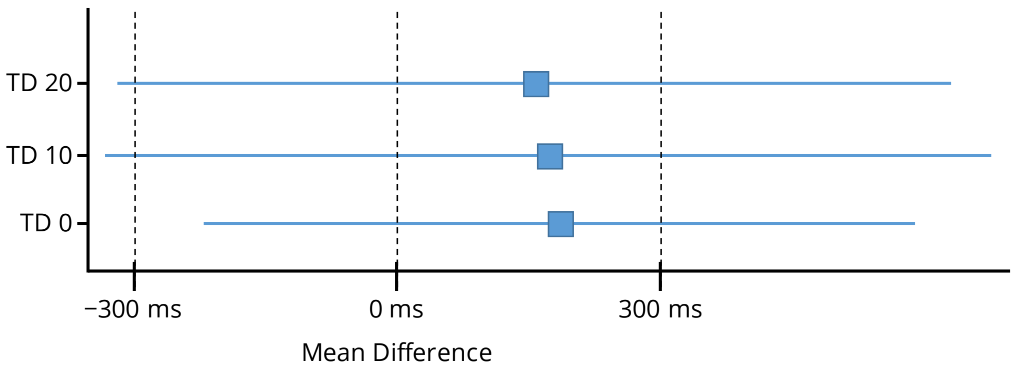

Figure 15. Illustration of the resulting confidence intervals for the mean difference in take-over times at different traffic densities.

times in this study? A statement of no influence of age would be premature not only because of the aforementioned methodological limitations but also because the data do not provide a conclusive answer: It is not possible to decide whether age influences take-over time by the data of this study alone. The recommendation here is to collect more data to arrive at a statistical decision.

9.4 Can every phenomenon be studied in a driving simulator? Obstacles and lessons learned in studying vigilance

The general notion that a passive monitoring role or long periods of conditionally automated driving lead to a vigilance decrement because of drowsiness has yet received mixed support in the literature (Feldhütter et al., 2018; Körber, Cingel et al., 2015; Schmidt et al., 2016). In Article 5, we investigated if the decrease in vigilance can be predicted by individual traits. To be precise, we investigated if a person's ability to stay vigilant and their boredom proneness is related to an individual's detection time of a malfunction of a driving automation system. The data showed no relationship between the predictors and the detection time. However, the general conclusion that there is no relationship between boredom proneness, vigilance ability, and detection time would be premature because of several reasons.

Firstly, partly due to the small sample size, the confidence intervals for the predictors are wide – there is low certainty about the estimators. Secondly, multiple confounding influences inherent in the operationalization could have washed out any existing relationship. Performance of sustained attention in partially automated driving was operationalized as the detection time of a suddenly occurring malfunction of lateral control. This malfunction was realized in an extensive right curve, where, without lateral control, the vehicle slowly drifted out of the lane. There are several factors that may conflate the relationship between vigilance level and reaction time. It is imaginable that the participants differed in their response bias, as known from signal detection theory (Stanislaw & Todorov, 1999), i.e. the individual threshold to classify the deviance from the lane center as a malfunction. In other words, a certain deviation from the lane center might be alarming for some participants while for others it may seem to be in line with a regularly functioning driving automation system. Secondly, different trust in automation may have led to differing degrees of reliance, leading to different thresholds for an intervention. As described in Article 4, lower trust leads to more voluntary interventions.

Thirdly, the current location of visual attention introduces a lot of error variance. Since the dependent variable was only measured at one point, there is no possibility to cancel out this random influence by performing multiple measurements. That might have drowned the relationship between the two predictors and the detection time in statistical noise.

Fourthly, the implemented vigilance task was a fast-paced task that was based on the resource theory of vigilance (Helton & Russell, 2011). According to this theory, detection failures in vigilance tasks can mainly be attributed to participants' depletion of their limited attentional resources. During a vigilance task, observers continuously make active discriminations between signal and noise with little to no opportunity for rest. By this, the limited attentional resources (Kahneman,

1973) deplete over time and a decline in performance follows. The stimulus rate of the vigilance task used in this study was set according to similar tasks to 50 trials per minute. It might very well be possible that this rate does not reflect the conditions that a driver encounters in partially automated driving. A typical highway drive in a partially automated vehicle lasts longer than the implemented task, but at the same time, the rate of stimuli (e.g., other road users) is probably lower. From this perspective, the vigilance test may not reflect the skills needed to stay vigilant during partially automated driving.

Fifthly, the experimental drive of an automated driving study is typically the first or one of the first drives in a driving simulator for the participants. The novelty of driving in a driving simulator and the situation of taking part in an experiment make it difficult to induce a state of monotony and boredom.

Taking these issues into account, the results of the study remain inconclusive on whether a decrement in vigilance can be predicted by the ability to sustain attention and boredom proneness. The recommendation regarding future research is to resolve the mentioned methodological problems. That means, for example, to implement a vigilance test that better reflects the activity partially automated driving. Also, a measure of vigilance has to be implemented that is less influenced by factors unrelated to vigilance such as an individual decision criterion, trust, or random error. Besides this, boredom proneness shows promise to play a meaningful role in automated driving apart from its potential effect on any performance measure. For example, studies on NRDT engagement could implement boredom proneness as a predictor of the onset and extent of engagement in an NDRT.

The abovementioned recommendation provokes the question if every mentioned methodological issue is solvable. Given the idiosyncrasies of driving simulators, mentioned in Section 9.2, the question arises if such a state of low vigilance can be elicited in them at all. No daylight and lack of risk but also the novelty of a driving simulation and the knowledge of being monitored may skew physiological processes. Thus, the individual ability to sustain attention is difficult to study in this setting. More generally, which kind of psychological subject can or cannot be studied in a driving simulator? A driving simulator provides controllability, reproducibility, and standardization of scenarios (e.g., traffic, weather conditions). They provide extended options for data recording, while measurements are accurate and synchronized. Their virtual environment allows the investigation of critical situations without negative consequences for the participants (de Winter, Leeuwen, & Happee, 2012). But therein also lies an essential drawback: Driving simulators lack real consequences and actions are free of any risk. This is not the case in a real drive and, thereby, a driving simulator exhibits limited ecological validity, especially in studies involving an NDRT (Dhami et al., 2004). All of the reported studies in this thesis have been conducted in a high-fidelity driving simulator, which provides a high level of immersion. The behavioral validity of such a simulator has been shown for many human factors research questions and even for take-over times (Eriksson, Banks, & Stanton, 2017). Although experiments on take-over time may exhibit certain external validity, the goal is not to draw conclusions about behavior in a simulator but behavior in real road traffic. Wickens et al. (2016) highlight the specific characteristic of data analysis in engineering psychology: Tightly controlled experiments such as those done in classic experimental psychology may fail to provide the right answers. This may be because effects found in such an environment are probably washed out when the effect is studied in a realistic applied situation where many other more or less random parameters influence behavior. Kimball (1957) defined these possibly incorrect conclusions as an "error of the third kind", "the error committed by giving the right answer to the wrong problem" (p. 134). Interaction with NDRTs, arousal, risk perception, or effects of long-term use may be completely different in naturalistic drives compared to simulator drives. For future studies, one has to first evaluate if the effect to be studied is significantly influenced by the difference between simulator and real drive. At the same time, current findings need to be replicated in naturalistic drives to estimate the ecologic validity of a driving simulator in such scenarios.

9.5 Challenges in the statistical analysis of individual differences

Just like a driving simulator cannot answer all research questions, it needs to be assessed if a one-size-fits-all data analysis approach might also not give us all the answers we seek. The currently dominant paradigm of data analysis in the domain of human factors is NHST, which is suitable and informative for certain research fields or questions but may not be optimal or even unsuitable for others (Szucs & Ioannidis, 2017). Its suitability for research on individual differences in automated driving is limited because in many cases it does not ask the questions that we seek to answer. The (mindless) null ritual consists of setting up a null hypothesis of no difference or zero correlation (nil hypothesis) without a specification of a substantive alternative hypothesis or any specific predictions. If the null hypothesis is rejected, the research hypothesis is accepted (Gigerenzer, 2004). However, NHST only allows rejection of the null hypothesis but tells nothing about how much the data fit a certain alternative hypothesis (Szucs & Ioannidis, 2017). In fact, an actual difference of exactly zero between two samples is the exception rather than the rule (Cohen, 1990), making a rejection of the null hypothesis the destruction of a mere strawman. By increasing precision or sample size (especially in large data sets), it is possible to reject almost any point hypothesis such as a hypothesis of exactly zero difference (Meehl, 1967). Based on the rationale to conduct applied research such as in human factors, the focus of engineering psychology does not lie on statistical significance but on practical significance. NHST may be unsuitable for this purpose because it tells nothing about the practical significance of an effect.

Contrarily, a focus on effect size estimation directly evaluates if an effect is large enough to cause worry instead of asking about its mere existence. The principles of applied research also have to be reflected in the design of the experiments of human factors studies. Laboratory settings are of limited value if the effects cannot be generalized to applied settings. In the same way, mindlessly maximizing fit of a model to a data set as a surrogate for a theory with parameters chosen post hoc to maximize the fit creates a model that accounts for almost everything after the fact but its ability to make new predictions is questionable (Gigerenzer, 1998). A successful fit to observed data has to be shown to not simply be due to a highly flexible model; the variability of the data and the likelihood of other outcomes has to be taken into account as well (Roberts & Pashler, 2000).

NHST is mostly accompanied by an analysis of mean differences. But as laid out in this thesis, individuals may strongly differ in their information perception, processing, or resulting behavior. Thus, summary statistics and statistical tests alone cannot answer if an individual interacts with automation in a safe manner – "no single index should substitute for scientific reasoning" (Wasserstein & Lazar, 2016, p. 132). Reliance on a dichotomous reject/non-reject null ritual results at best in an incomplete understanding of human-automation interaction. Besides this, safety does not have to be ensured solely for the average user but also needs the examination of the full variety of participant behavior, distributions, outliers, and rare events. Rare events must not be treated as outliers but need to be investigated carefully. While the interaction with two different systems might not manifest in mean differences, the frequency of rare events could differ. The research of drowsiness in automated driving has further shown that drowsiness may develop in an automated drive but its onset and extent is highly variable. This thesis also showed that individual differences among drivers can decisively influence their interaction with automation. For this reason, a reliance on solely the sample mean may be misleading and uninformative to estimate the safety of automated driving. Estes (1956) reconciles both approaches in his recommendation: "The group curve will remain one of our most useful devices both for summarizing information and for theoretical analysis provided only that it is handled with a modicum of tact and understanding" (p. 134). All things considered, neither a focus on averages or groups nor a completely individual analysis of each participant is a viable analysis strategy. Both each answer research questions on their own and both should be consulted to answer human factors research questions.

The recommendation derived from the results of this thesis is to accommodate the data analysis procedures to allow for an analysis of individual differences as well. That means to implement a holistic analysis using a toolbox instead of mindless one-size-fits-all analysis by the book. This can be accomplished by considering the whole distribution of observations – the central tendency (e.g., mean), extreme or rare values, percentiles and individual development in longitudinal designs. Dichotomous NHST decision may be substituted with effect size estimation combined with expressing uncertainty by interval estimation (Cumming, 2014; Kruschke & Liddell, 2018), quantification of evidence (Dienes & Mclatchie, 2018; Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011), a strict distinction between exploratory and confirmatory research (Wagenmakers et al., 2012), and a focus on practical relevance/significance (Radlmayr, Körber, Feldhütter, & Bengler, 2016). A further discussion is provided in the complementary articles in Section A3 of the appendix.

9.6 Users adapt to automated driving (to a certain extent)

Adaption and learning are well-known and fundamental psychological processes that appear within every experience and driving is no exception (Bubb et al., 2015). Consequently, with growing experience, drivers get better at driving (McCartt, Mayhew, Braitman, Ferguson, & Simpson, 2009). This is also visible in take-over data: In Article 2, the take-over times decreased from the 3.70 s in the first take-over situation to 2.80 s in the last take-over situation. But the drivers do not only learn to perform a take-over, they also adapt their multitasking behavior. Article 3 contained five take-over situations where the participants were engaged in an NDRT while driving conditionally automated. In the course of the five situations, the mean take-over time decreased from 737.17 to −1687.53 ms (i.e. before the TOR was even issued).

However, this adaption seems to have limits. We split the sample into four quartiles depending their multitasking ability. After experiencing two take-over situations, the means of the first, second, and third quartile do not differ anymore; the drivers adapted to the multitasking situation. The fourth quartile, the worst multitaskers, still remained at a constant mean difference of about 1689 ms in take-over time to the other three quartiles throughout all take-over situations. Thus, all drivers reacted faster, but a stable performance difference still persisted. The results are consistent with an adaption in NDRT task management strategy but also highlight that even after changing the strategy, stable individual differences in the ability remained. Thus, we need to realize and then work on two things. First, the results show that drivers started with different strategies for engaging in the NDRT. Future studies need to identify predictors of the choice of strategy, such as risk preference, trust in automation, or the presentation of an automated system. Second, even if certain drivers adapt their strategy, they do not reach the same multi-tasking performance as others. Future work needs to reinforce the evidence and to work out other potential stable performance differences as well as their variation according to different contexts.

Albeit the correlation coefficients between the multi-tasking test results and take-over time decrease after the second take-over situation, they are still at a very high level compared to other relationships in individual differences research. Studies on individual difference generally do not obtain effect sizes as large as other subdomains of automated driving research such as HMI design. For example, Gignac and Szodorai (2016) derived effect size guidelines for individual difference research based on meta-analytically acquired correlations and suggest considering a correlation of only $r = .29$ as the benchmark for a large effect. At the same time, small samples are more vulnerable to over-fit (Gelman & Carlin, 2014). The study, therefore, provides substantive initial evidence, still, it should be replicated with a larger sample to obtain more precise effect size estimates. Likewise, the use of a more modern data analysis method such as a mixed-effect model is recommended (Bates, Mächler, Bolker, & Walker, 2015). While the implemented multitasking test exhibits a high face validity, the underlying mechanisms causing participants to differ in their test performance can yet only be hypothesized (see Section 6.1). The recommendation regarding future research here is to not only replicate the finding but also to position it in the current theories on multitasking performance by deriving and testing hypothesized causal mechanisms.

What is more, Article 3 highlights that switch costs as a result of multitasking inhere a substantial potential for the design of the take-over process. The study's results suggest some recommendations on how the design of an HMI and TORs may facilitate the take-over process by reducing switch costs and by supporting the shift of attention. Individuals are generally better at detecting and reacting to an object if they know in advance something about its features, such as its location or motion (Corbetta & Shulman, 2002; Dosher & Lu, 2000; Posner et al., 1980). Based on this, directional warnings can prompt an overt or covert shift of attention towards a hazard, for example by emitting the TOR from the hazard's direction. Indeed, directed warning signals have already been shown to reduce reaction times to hazards (Weller et al., 2011). A cue could also provide an evident indication of what action should be performed since a response to a stimulus is quicker when a participant knows in advance what movement they have to make (Abrams & Jonides, 1988; Allport, Styles, & Hsieh, 1994; Arbuthnott & Woodward, 2002; Corbetta & Shulman, 2002; Rosenbaum, 1980). This approach could also be combined with a directional cue: The Simon effect (Simon & Rudell, 1967) describes that the reaction time to a stimulus is lower if the location of the stimulus corresponds to the location of the required response (VandenBos, 2015).

9.7 Users (could) maladapt to automated driving

The abovementioned study has shown that drivers learn and adapt in the interaction with automation. Consequently, the take-over time reduces with repeated experience of a take-over situation (Körber, Gold et al., 2016; Körber, Weißgerber et al., 2015). But this does not reflect the complete big picture: Human-automation interaction represents the paradoxical situation that users may deliver worse performance with increasing experience. Behavioral adaption is already known from studies on ADAS, in particular on ACC (Hoedemaeker & Brookhuis, 1998; Mahr & Müller, 2011; Rudin-Brown & Parker, 2004). While participants behave according to instructions in single drives in experiments, long-term use can drastically differ in terms of risky behavior and rule obedience (Mahr & Müller, 2011). Many current studies investigate the interaction with a driving automation system only in a single drive. But complacent behavior does typically not come up within the first interactions but after prolonged exposure to a reliable system (Parasuraman & Manzey, 2010). The usual development of trust reflects a steady increase until a certain high and stable level (Beggiato et al., 2015). If a system performs on a very high level of reliability, users could build up over-reliance and interact with automation in ways not intended by design. High reliability may lead to complacency, as mentioned in Section 7.4, which describes reduced information sampling and verification of the automated system's decisions and recommendations (Bahner et al., 2008; Parasuraman & Manzey, 2010).

Accordingly, Article 4 has shown that higher levels of trust are associated with less monitoring and lower take-over performance. There is also already initial anecdotal evidence for this process in automated driving: Brown, the aforementioned Tesla driver whose accident was narrated at the beginning, built up so much trust in Autopilot ("The car's doing it all itself") that he allegedly started to watch a movie when he was actually responsible for monitoring the system. Likewise, false alarms may promote maladaptive behavioral adaption: A high number of false alarms reduce compliance to alarms and lead to a *cry wolf* effect where the user starts to ignore warnings or at least reacts to them very slowly, for example after finishing another task (Dixon & Wickens, 2006).

The results of Article 3 provide empirical support for this notion of behavioral adaption to a driving automat system, as well as the evolution and development of strategies for attention allocation. Brown's accident provides anecdotal evidence for its potential hazardous consequences. The task is now to investigate how drivers adapt to automated driving in long-term day-to-day use. The recommendation is to widen the focus from cross-sectional studies to longitudinal studies to investigate drivers' long-term adaption to a driving automation system. This could be realized either in repeated experimental drives, long-term field operational tests, or by interviews and questionnaires on usage patterns.

9.8 Automation is neither good nor bad in itself

Implementing automation in itself is neither a good nor a bad decision – the deciding factor is how it is implemented. The introduction of automation in other domains has shown that the economic and safety benefits are often accompanied by novel problems. Automation can enhance performance and make us smart, or degrade performance and make us dumb (Christoffersen & Woods, 2002; Norman, 2003). A higher degree of automation is not simply more of the same; each degree represents a human-machine system on its own, reflecting a qualitative change in the interaction (Bengler, Winner, & Wachenfeld, 2017). Human and automated system are not independent: Implementing automation alters the quality of the interaction, the operator's role and responsibilities, and the nature of the cognitive demands. To argue that we can overcome these novel problems with simply more automation ignores the fact that technology cannot be considered independent of its human users (Christoffersen & Woods, 2002; Sarter et al., 1997). Hence, a driver-centered perspective is a key requirement for the introduction of automated vehicles. Implementation of a driver-centered design approach requires an understanding of the capabilities, vulnerabilities, and idiosyncrasies of both driver and driving automation system in the context of different tasks and situations (Drnec et al., 2016). This issue is especially pronounced in automated driving, where future users are not a homogenous group of screened and trained experts but a heterogeneous group of laypeople.

The results of this thesis empirically support this point of view: The studies systematically revealed that individual differences among drivers can have a safety-relevant influence on their interaction with a driving automation system. This highlights the importance of a driver-centered perspective in the design of automated vehicles. The driver-centered design process of vehicles must now recognize the safety-relevant influence of individual differences and, consequently, it must also comprise the design of the driving automation system. Articles 3 and 4 demonstrated that individual differences in trust and multitasking ability can crucially influence drivers' allocation of attention and their reactions to critical events, which results in observable and serious consequences. This thesis hereby provides empirical support the recommendation to enrich technological considerations with driver-centered design. The implied recommendation is a cautious evidence-based development strategy that takes the drivers with all of their capabilities, vulnerabilities, and idiosyncrasies into account. Vice versa, in light of the findings of this thesis, a development process that solely relies on technical reliability may not suffice to ensure a safe introduction of automated vehicles to road traffic.

The investigation of individual differences remains an important topic. First, because they can have serious consequences and, second, because it is difficult to deduce which individual difference really translates into such consequences. Article 2, despite our expectations, did not provide sufficient evidence for an effect of age on take-over performance. Not every difference among drivers translates into an observable difference in their interaction with automation. This provides further support for a step-by-step evidence-based development process.

As laid out in this discussion, the investigation of individual differences and drawing valid conclusions is challenging. The studies conducted as part of this thesis investigated individual differences in specific situations with a focus on the take-over process in a driving simulator. The aforementioned constraints imposed by the research setting somewhat limited the extent of the progress that has been possible in the scope of this thesis. Yet, when combined, these small steps led to a better understanding of individual differences in the interaction with automation. As Cy Levinthal once remarked:

> Well, there are two kinds of biologists, those who are looking to see if there is one thing that can be understood, and those who keep saying it is very complicated and that nothing can be understood. . . . You must study the simplest system you think has the properties you are interested in. (Platt, 1964, pp. 348–349)

References

Abe, G., Itoh, M., & Tanaka, K. (2002). Dynamics of drivers' trust in warning systems. In L. Basañez (Ed.): *IFAC proceedings volumes, Proceedings of the 15th triennial world congress of the International Federation of Automatic Control* (pp. 363–368). New York, NY: Elsevier. https://doi.org/10.3182/20020721-6-ES-1901.01614

Abendroth, B., & Bruder, R. (2016). Capabilities of humans for vehicle guidance. In H. Winner, S. Hakuli, F. Lotz, & C. Singer (Eds.), *Handbook of driver assistance systems: Basic information, components and systems for active safety and comfort* (pp. 3–18). Cham: Springer Reference.

Abraham, H., Reimer, B., Seppelt, B. D., Fitzgerald, C., Mehler, B., & Coughlin, J. F. (2018). Consumer interest in automation: Change over one year. In *Proceedings of The Transportation Research Board 97th Annual Meeting.*

Abrams, R. A., & Jonides, J. (1988). Programming saccadic eye movements. *Journal of Experimental Psychology. Human Perception and Performance*, *14*(3), 428–443.

Adell, E., Várhelyi, A., & Nilsson, L. (2014). The definition of acceptance and acceptability. In M. A. Regan, T. Horberry, & A. Stevens (Eds.), *Human factors in road and rail transport. Driver acceptance of new technology: Theory, measurement and optimisation* (pp. 11–21). Farnham, Surrey, England, UK: Ashgate Publishing Ltd.

Aeberhard, M., Rauch, S., Bahram, M., Tanzmeister, G., Thomas, J., Pilat, Y., . . . Kämpchen, N. (2015). Experience, results and lessons learned from automated driving on Germany's highways. *IEEE Intelligent Transportation Systems Magazine*, *7*(1), 42–57. https://doi.org/10.1109/MITS.2014.2360306

Ahmed, N., de Visser, E., Shaw, T., Mohamed-Ameen, A., Campbell, M., & Parasuraman, R. (2014). Statistical modelling of networked human-automation performance using working memory capacity. *Ergonomics*, *57*(3), 295–318. https://doi.org/10.1080/00140139.2013.855823

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior.* Englewood Cliffs, NJ: Prentice Hall.

Allahyari, T., Saraji, G. N., Adl, J., Hosseini, M., Iravani, M., Younesian, M., & Kass, S. J. (2008). Cognitive failures, driving errors and driving accidents. *International Journal of Occupational Safety and Ergonomics*, *14*(2), 149–158. https://doi.org/10.1080/10803548.2008.11076759

Allen, B. P., & Potkay, C. R. (1981). On the arbitrary distinction between states and traits. *Journal of Personality and Social Psychology*, *41*(5), 916–928. https://doi.org/10.1037/0022-3514.41.5.916

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. A. Umiltà & M. Moscovitch (Eds.), *Attention and performance: Vol. 15. Conscious and nonconscious information processing* (pp. 421–452). Cambridge, Mass: MIT Press.

Alzahabi, R., & Becker, M. W. (2013). The association between media multitasking, task-switching, and dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1485–1495. https://doi.org/10.1037/a0031208

Amelang, M., & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention* [Psychological diagnostics and intervention] (4th ed.). *Springer-Lehrbuch*. Berlin, Heidelberg: Springer Medizin.

Anderson, P. J., Vicki, A., & Jacobs, R. (Eds.). (2014). *Executive functions and the frontal lobes: A lifespan perspective* (1st ed.). *Studies on neuropsychology, neurology, and cognition*. New York: Routledge.

Andrews, E. C., & Westerman, S. J. (2012). Age differences in simulated driving performance: Compensatory processes. *Accident Analysis & Prevention*, *45*, 660–668. https://doi.org/10.1016/j.aap.2011.09.047

Annett, J. (2002). Subjective rating scales: Science or art? *Ergonomics*, *45*(14), 966–987. https://doi.org/10.1080/00140130210166951

Anstey, K. J., Wood, J., Lord, S., & Walker, J. G. (2005). Cognitive, sensory and physical factors enabling driving safety in older adults. *Clinical Psychology Review*, *25*(1), 45–65. https://doi.org/10.1016/j.cpr.2004.07.008

Arbuthnott, K. D., & Woodward, T. S. (2002). The influence of cue-task association and location on switch cost and alternating-switch cost. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale*, *56*(1), 18–29. https://doi.org/10.1037/h0087382

Arthur, W., Barrett, G. V., & Alexander, R. A. (1991). Prediction of vehicular accident involvement: A meta-analysis. *Human Performance*, *4*(2), 89–105. https://doi.org/10.1207/s15327043hup0402_1

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, *2*, 89–195.

Baddeley, A. D. (2007). *Working memory, thought, and action. Oxford psychology series: Vol. 45*. Oxford: Oxford University Press.

Bagheri, N., & Jamieson, G. A. (2004). Considering subjective trust and monitoring behavior in assessing automation-induced "complacency". In D. A. Vincenzi, M. Mouloua, & P. A. Hancock (Eds.), *Human performance, situation awareness, and automation: Current research and trends HPSAA II, Volumes I and II* (Vol. 1, pp. 54–59). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Bagozzi, R. P. (1982). The role of measurement in theory construction and hypothesis testing: Toward a holistic model. In C. Fornell (Ed.), *Praeger scientific. A second generation of multivariate analysis* (pp. 5–23). New York, NY: Praeger.

Bahner, J. E., Hüper, A.-D., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies*, *66*(9), 688–699. https://doi.org/10.1016/j.ijhcs.2008.06.001

Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*, *8*(4), 321–348. https://doi.org/10.1080/14639220500535301

Baldock, M. R. J., Mathias, J. L., McLean, A. J., & Berndt, A. (2006). Self-regulation of driving and its relationship to driving ability among older adults. *Accident Analysis & Prevention*, *38*(5), 1038–1045. https://doi.org/10.1016/j.aap.2006.04.016

Baldwin, C. L., & Schieber, F. (1995). Dual task assessment of age differences in mental workload with implications for driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *39*(2), 167–171. https://doi.org/10.1177/154193129503900210

Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology*, *50*(1), 471–507. https://doi.org/10.1146/annurev.psych.50.1.471

Barber, B. (1983). *The logic and limits of trust*. New Brunswick, N.J.: Rutgers University Press.

Barg-Walkow, L. H., & Rogers, W. A. (2016). The effect of incorrect reliability information on expectations, perceptions, and use of automation. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *58*(2), 242–260. https://doi.org/10.1177/0018720815610271

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Baumann, M., Rösler, D., & Krems, J. F. (2007). Situation awareness and secondary task performance while driving. In D. Harris (Ed.), *Lecture notes in computer science. Engineering psychology and cognitive ergonomics* (Vol. 4562, pp. 256–263). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73331-7_27

Beggiato, M., & Krems, J. F. (2013). The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information. *Transportation Research Part F: Traffic Psychology and Behaviour*, *18*, 47–57. https://doi.org/10.1016/j.trf.2012.12.006

Beggiato, M., Pereira, M., Petzoldt, T., & Krems, J. F. (2015). Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *35*, 75–84. https://doi.org/10.1016/j.trf.2015.10.005

Bélanger, A., Gagnon, S., & Stinchcombe, A. (2015). Crash avoidance in response to challenging driving events: The roles of age, serialization, and driving simulator platform. *Accident Analysis & Prevention*, *82*, 199–212. https://doi.org/10.1016/j.aap.2015.04.030

Beller, J., Heesen, M., & Vollrath, M. (2013). Improving the driver-automation interaction: An approach using automation uncertainty. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *55*(6), 1130–1141. https://doi.org/10.1177/0018720813482327

Bengler, K., Dietmayer, K., Färber, B., Maurer, M., Stiller, C., & Winner, H. (2014). Three decades of driver assistance systems: Review and future perspectives. *IEEE Intelligent Transportation Systems Magazine*, *6*(4), 6–22. https://doi.org/10.1109/MITS.2014.2336271

Bengler, K., Winner, H., & Wachenfeld, W. (2017). No human – No cry? *At - Automatisierungstechnik*, *65*(7). https://doi.org/10.1515/auto-2017-0021

Bengler, K., Zimmermann, M., Bortot, D., Kienle, M., & Damböck, D. (2012). Interaction principles for cooperative human-machine systems. *It - Information Technology*, *54*(4), 157–164. https://doi.org/10.1524/itit.2012.0680

Van den Beukel, A. P., van der Voort, M. C., & Eger, A. O. (2016). Supporting the changing driver's task: Exploration of interface designs for supervision and intervention in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *43*, 279–301. https://doi.org/10.1016/j.trf.2016.09.009

Billings, C. E. (1997). *Aviation automation: The search for a human-centered approach*. Mahwah, N.J.: Lawrence Erlbaum Associates Publishers.

Bisantz, A. M., & Seong, Y. (2001). Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics*, *28*(2), 85–97. https://doi.org/10.1016/S0169-8141(01)00015-4

Blackwelder, W. C. (1982). "Proving the null hypothesis" in clinical trials. *Controlled Clinical Trials*, *3*(4), 345–353. https://doi.org/10.1016/0197-2456(82)90024-1

Blana, E. (1996). *Driving simulator validation studies: A literature review* (Working Paper No. 480). Leeds, UK.

Blomqvist, K. (1997). The many faces of trust. *Scandinavian Journal of Management*, *13*(3), 271–286. https://doi.org/10.1016/S0956-5221(97)84644-1

Borst, J. P., Taatgen, N. A., & van Rijn, H. (2010). The problem state: A cognitive bottleneck in multitasking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 363–382. https://doi.org/10.1037/a0018106

Boudette, N. E. (2017, January 19). Tesla's self-driving system cleared in deadly crash. *The New York Times*. Retrieved from https://www.nytimes.com/2017/01/19/business/tesla-model-s-autopilot-fatal-crash.html

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & G. N. Wilkinson (Eds.), *Robustness in statistics* (pp. 201–236). New York: Academic Press. https://doi.org/10.1016/B978-0-12-438150-6.50018-2

Braitman, K. A., Kirley, B. B., Ferguson, S., & Chaudhary, N. K. (2007). Factors leading to older drivers' intersection crashes. *Traffic Injury Prevention*, *8*(3), 267–274. https://doi.org/10.1080/15389580701272346

Brewer, M. B., & Crano, W. D. (2014). Research design and issues of validity. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 11–26). New York, NY: Cambridge University Press.

Breznitz, S. (1984). *Cry wolf: The psychology of false alarms*. Hillsdale, N.J.: Lawrence Erlbaum Associates Publishers.

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.

Broughton, J., & Markey, K. A. (1996). *In-car equipment to help drivers avoid accidents* (TRL No. 198). Crowthorne.

Brown, J. (2015). *Tesla Autopilot v7.0 stop and go traffic*. Retrieved from https://www.youtube.com/watch?v=5TjbqVartjM

Brown, R. D., & Galster, S. M. (2004). Effects of reliable and unreliable automation on subjective measures of mental workload, situation awareness, trust and confidence in a dynamic flight task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2004* (pp. 147–151).

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, Los Angeles, CA: University of California Press.

Bryan, J., & Luszcz, M. A. (2000). Measurement of executive function: Considerations for detecting adult age differences. *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section a)*, *22*(1), 40–55. https://doi.org/10.1076/1380-3395(200002)22:1;1-8;ft040

Bubb, H. (1992). *Menschliche Zuverlässigkeit: Definitionen, Zusammenhänge, Bewertung* [Human reliability: Definitions, relationships, evaluation] (1st ed.). Landsberg: Ecomed.

Bubb, H. (2015). Einführung [Introduction]. In H. Bubb, K. Bengler, R. E. Grünen, & M. Vollrath (Eds.), *Automobilergonomie* (pp. 1–25). Wiesbaden: Springer Fachmedien Wiesbaden.

Bubb, H., Bengler, K., Grünen, R. E., & Vollrath, M. (Eds.). (2015). *Automobilergonomie* [Automotive ergonomics]. Wiesbaden: Springer Fachmedien Wiesbaden.

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* [Introduction to test and questionnaire construction] (3rd ed.). *Psychologie*. München: Pearson Studium.

Bühner, M., König, C. J., Pick, M., & Krumm, S. (2006). Working memory dimensions as differential predictors of the speed and error aspect of multitasking performance. *Human Performance*, *19*(3), 253–275. https://doi.org/10.1207/s15327043hup1903_4

Bukasa, B., & Utzelmann, H. D. (2009). Psychologische Diagnostik der Fahreignung [Psychological diagnostics of the fitness to drive]. In H.-P. Krüger & N. Birbaumer (Eds.), *Enzyklopädie der Psychologie Praxisgebiete Verkehrspsychologie: Bd. 2. Anwendungsfelder der Verkehrspsychologie* (pp. 237–275). Göttingen: Hogrefe Verlag für Psychologie.

Bundesministerium für Verkehr und digitale Infrastruktur. (2015). Strategie automatisiertes und vernetztes Fahren [Strategy automated and connected driving]. Retrieved from https://www.bmvi.de/SharedDocs/DE/Publikationen/StB/broschuere-strategie-automatisiertes-vernetztes-fahren.pdf?__blob=publicationFile

Burisch, M. (1978). Construction strategies for multiscale personality inventories. *Applied Psychological Measurement*, *2*(1), 97–111. https://doi.org/10.1177/014662167800200110

Burisch, M. (1984). Approaches to personality inventory construction: A comparison of merits. *American Psychologist*, *39*(3), 214–227. https://doi.org/10.1037/0003-066X.39.3.214

Burns, P. C., & Wilde, G. J. S. (1995). Risk taking in male taxi drivers: Relationships among personality, observational data and driver records. *Personality and Individual Differences*, *18*(2), 267–278. https://doi.org/10.1016/0191-8869(94)00150-q

Butler, J. K. (1991). Toward understanding and measuring conditions of trust: Evolution of a conditions of trust inventory. *Journal of Management*, *17*(3), 643–663. https://doi.org/10.1177/014920639101700307

Butler, J. K., & Cantrell, R. S. (1984). A behavioral decision theory approach to modeling dyadic trust in superiors and subordinates. *Psychological Reports*, *55*(1), 19–28. https://doi.org/10.2466/pr0.1984.55.1.19

Caggiano, D. M., & Parasuraman, R. (2004). The role of memory representation in the vigilance decrement. *Psychonomic Bulletin & Review*, *11*(5), 932–937. https://doi.org/10.3758/BF03196724

Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*, *3*(3), 106–116. https://doi.org/10.3844/jssp.2007.106.116

Carlson, M. S., Drury, J. L., Desai, M., Kwak, H., & Yanco, H. A. (2014). Identifying factors that influence trust in automated cars and medical diagnosis systems. In *2014 AAAI Spring Symposium Series: The intersection of robust intelligence and trust in autonomous systems* (pp. 20–27). Palo Alto, California.

Casner, S. M., Hutchins, E. L., & Norman, D. A. (2016). The challenges of partially automated driving. *Communications of the ACM*, *59*(5), 70–77. https://doi.org/10.1145/2830565

Cauley, J. A. (2012). The demography of aging. In A. B. Newman & J. A. Cauley (Eds.), *The epidemiology of aging* (pp. 3–14). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-5061-6_1

Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of states and traits: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology*, *54*(4), 541–557.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press. https://doi.org/10.1016/B978-0-12-170150-5.50011-1

Choi, J. K., & Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction*, *31*(10), 692–702. https://doi.org/10.1080/10447318.2015.1070549

Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research: Vol. 2. Automation* (1st ed., pp. 1–12). Amsterdam, Boston: JAI. https://doi.org/10.1016/S1479-3601(02)02003-9

Churchill, G. A., & Peter, J. P. (1984). Research design effects on the reliability of rating scales: A meta-analysis. *Journal of Marketing Research*, *21*(4), 360. https://doi.org/10.2307/3151463

Clark, H., & Feng, J. (2017). Age differences in the takeover of vehicle control and engagement in non-driving-related activities in simulated driving with conditional automation. *Accident Analysis & Prevention*, *106*, 468–479. https://doi.org/10.1016/j.aap.2016.08.027

Clarke, D. D., Ward, P., Bartle, C., & Truman, W. (2010). Older drivers' road traffic crashes in the UK. *Accident Analysis & Prevention*, *42*(4), 1018–1024. https://doi.org/10.1016/j.aap.2009.12.005

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*(12), 1304–1312. https://doi.org/10.1037/0003-066X.45.12.1304

Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *The Journal of Applied Psychology*, *92*(4), 909–927. https://doi.org/10.1037/0021-9010.92.4.909

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Creaser, J. I., & Fitch, G. M. (2015). Human factors considerations for the design of Level 2 and Level 3 automated vehicles. In G. Meyer & S. Beiker (Eds.), *Lecture notes in mobility. Road vehicle automation 2* (pp. 81–89). Cham: Springer International Publishing.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671–684. https://doi.org/10.1037/h0043943

Crundall, D., Underwood, G., & Chapman, P. (1999). Driving experience and the functional field of view. *Perception*, *28*(9), 1075–1087. https://doi.org/10.1068/p281075

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Czaja, S. J., Charness, N., Fisk, A. D., Hertzog, C., Nair, S. N., Rogers, W. A., & Sharit, J. (2006). Factors predicting the use of technology: Findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). *Psychology and Aging*, *21*(2), 333–352. https://doi.org/10.1037/0882-7974.21.2.333

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *351*(1346), 1413–1420. https://doi.org/10.1098/rstb.1996.0125

Das, S., Sun, X., Wang, F., & Leboeuf, C. (2015). Estimating likelihood of future crashes for crash-prone drivers. *Journal of Traffic and Transportation Engineering (English Edition)*, *2*(3), 145–157. https://doi.org/10.1016/j.jtte.2015.03.003

Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., . . . Starr, J. M. (2009). Age-associated cognitive decline. *British Medical Bulletin*, *92*, 135–152. https://doi.org/10.1093/bmb/ldp033

Deary, I. J., & Der, G. (2005). Reaction time, age, and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples. *Aging, Neuropsychology, and Cognition (Neuropsychology, Development and Cognition: Section B)*, *12*(2), 187–215. https://doi.org/10.1080/13825580590969235

Deery, H. A. (1999). Hazard and risk perception among young novice drivers. *Journal of Safety Research*, *30*(4), 225–236. https://doi.org/10.1016/S0022-4375(99)00018-3

DeJoy, D. M. (1992). An examination of gender differences in traffic accident risk perception. *Accident Analysis & Prevention*, *24*(3), 237–246. https://doi.org/10.1016/0001-4575(92)90003-2

Der, G., & Deary, I. J. (2006). Age and sex differences in reaction time in adulthood: Results from the United Kingdom Health and Lifestyle Survey. *Psychology and Aging*, *21*(1), 62–73. https://doi.org/10.1037/0882-7974.21.1.62

Deutsch, M. (1958). Trust and suspicion. *Journal of Conflict Resolution*, *2*(4), 265–279. https://doi.org/10.1177/002200275800200401

Deutsch, M. (1960). The effect of motivational orientation upon trust and suspicion. *Human Relations*, *13*(2), 123–139. https://doi.org/10.1177/001872676001300202

DeVellis, R. F. (2006). Classical test theory. *Medical Care*, *44*(11), 50-59. https://doi.org/10.1097/01.mlr.0000245426.10853.30

Devlin, A., McGillivray, J., Charlton, J., Lowndes, G., & Etienne, V. (2012). Investigating driving behaviour of older drivers with mild cognitive impairment using a portable driving simulator. *Accident Analysis & Prevention*, *49*, 300–307. https://doi.org/10.1016/j.aap.2012.02.022

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, *130*(6), 959–988. https://doi.org/10.1037/0033-2909.130.6.959

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: A predictive validity perspective. *Journal of the Academy of Marketing Science*, *40*(3), 434–449. https://doi.org/10.1007/s11747-011-0300-3

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review*, *25*(1), 207–218. https://doi.org/10.3758/s13423-017-1266-z

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Basingstoke, Hampshire: Palgrave Macmillan.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(10), 2636–2641. https://doi.org/10.1073/pnas.1513271113

Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *48*(3), 474–486. https://doi.org/10.1518/001872006778606822

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431.

Donmez, B., Boyle, L. N., Lee, J. D., & McGehee, D. V. (2006). Drivers' attitudes toward imperfect distraction mitigation strategies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *9*(6), 387–398. https://doi.org/10.1016/j.trf.2006.02.001

Dosher, B. A., & Lu, Z. L. (2000). Mechanisms of perceptual attention in precuing of location. *Vision Research*, *40*(10-12), 1269–1292.

Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 years after Likert: Thurstone was right! *Industrial and Organizational Psychology*, *3*(04), 465–476. https://doi.org/10.1111/j.1754-9434.2010.01273.x

Drnec, K., Marathe, A. R., Lukos, J. R., & Metcalfe, J. S. (2016). From trust in automation to decision neuroscience: Applying cognitive neuroscience methods to understand and improve interaction decisions involved in human automation interaction. *Frontiers in Human Neuroscience*, *10*, 54. https://doi.org/10.3389/fnhum.2016.00290

Dzindolet, M. T., Beck, H. P., Pierce, L. G., & Dawe, L. A. (2001). *A framework of automation use* (No. ARL-TR-2412). Aberdeen Proving Ground, MD.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, *58*(6), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *44*(1), 79–94.

Eisinga, R., Grotenhuis, M. t., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Ellis, D., & Allaire, J. C. (1999). Modeling computer interest in older adults: The role of age, education, computer knowledge, and computer anxiety. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *41*(3), 345–355. https://doi.org/10.1518/001872099779610996

Elson, M. (2017). Question wording and item formulation. In J. Matthes, R. Potter, & C. S. Davis (Eds.), *Wiley Blackwell-ICA International Encyclopedias of Communication. The international encyclopedia of communication research methods.* Hoboke, New Jersey: Wiley-Blackwell.

Emerson, J. L., Johnson, A. M., Dawson, J. D., Uc, E. Y., Anderson, S. W., & Rizzo, M. (2012). Predictors of driving outcomes in advancing age. *Psychology and Aging*, *27*(3), 550–559. https://doi.org/10.1037/a0026359

Emons, W. H. M., Sijtsma, K., & Meijer, R. R. (2007). On the consistency of individual classification using short scales. *Psychological Methods*, *12*(1), 105–120. https://doi.org/10.1037/1082-989X.12.1.105

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *32*(2), 97–101. https://doi.org/10.1177/154193128803200221

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *37*(2), 381–394. https://doi.org/10.1518/001872095779064555

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Engström, J. (2011). Understanding attention selection in driving: From limited capacity to adaptive behaviour (Doctoral dissertation). Chalmers University of Technology, Göteborg.

Eriksson, A., Banks, V. A., & Stanton, N. A. (2017). Transition to manual: Comparing simulator with on-road control transitions. *Accident Analysis & Prevention*, *102*, 227–234. https://doi.org/10.1016/j.aap.2017.03.011

Eskandarian, A. (Ed.). (2012). *Handbook of intelligent vehicles. Springer reference*. London: Springer.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*(2), 134–140. https://doi.org/10.1037/h0045156

European Commission. (2015). *Road safety in the European Union: Trends, statistics and main challenges, March 2015*. Brüssel.

European Commission. (2016). *2015 road safety statistics: What is behind the figures?* (MEMO No. 16/864). Brüssel. Retrieved from European Commission website: http://europa.eu/rapid/press-release_MEMO-16-864_en.htm

Evans, L., & Wasielewski, P. (1983). Risky driving related to driver and vehicle characteristics. *Accident Analysis & Prevention*, *15*(2), 121–136. https://doi.org/10.1016/0001-4575(83)90068-4

Eysenck, M. W., & Keane, M. T. (2010). *Cognitive psychology: A student's handbook* (6th ed.). Hove: Psychology Press.

Farmer, R., & Sundberg, N. D. (1986). Boredom proneness – The development and correlates of a new scale. *Journal of Personality Assessment*, *50*(1), 4–17. https://doi.org/10.1207/s15327752jpa5001_2

Feldhütter, A., Gold, C., Hüger, A., & Bengler, K. (2016). Trust in automation as a matter of media influence and experience of automated vehicles. In *Proceedings of the Human Factors and Ergonomics Society (HFES) 60th Annual Meeting 2016* (Vol. 60, pp. 2024–2028). https://doi.org/10.1177/1541931213601460

Feldhütter, A., Hecht, T., Kalb, L., & Bengler, K. (2018). Effect of prolonged periods of conditionally automated driving on the development of fatigue: With and without non-driving-related activities. *Cognition, Technology & Work*, *12*(1), 1–8. https://doi.org/10.1007/s10111-018-0524-9

Feuerberg, B. V., Bahner, J. E., & Manzey, D. (2005). Interindividuelle Unterschiede im Umgang mit Automation – Entwicklung eines Fragebogens zur Erfassung des Complacency-Potentials [Interindividual differences in the interaction with automation – Development of a questionnaire to assess potential for complacency]. In L. Urbas & C. Steffens (Eds.), *Zustandserkennung und Systemgestaltung. 6. Berliner Werkstatt Mensch-Maschine-Systeme.* (pp. 199–202). Düsseldorf: VDI-Verlag.

Fiedler, K. (2011). Voodoo correlations are everywhere–not only in neuroscience. *Perspectives on Psychological Science*, *6*(2), 163–171. https://doi.org/10.1177/1745691611400237

Finn, P., & Bragg, B. W. (1986). Perception of the risk of an accident by young and older drivers. *Accident Analysis & Prevention*, *18*(4), 289–298.

Finomore, V. S., Matthews, G., Shaw, T. H., & Warm, J. S. (2009). Predicting vigilance: A fresh look at an old problem. *Ergonomics*, *52*(7), 791–808. https://doi.org/10.1080/00140130802641627

Fisk, A. D., Ackerman, P. L., & Schneider, W. (1987). Automatic and controlled processing theory and its applications to human factors problems. In P. A. Hancock (Ed.), *Advances in psychology: Vol. 47. Human Factors Psychology* (Vol. 47, pp. 159–197). Amsterdam: Elsevier.

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research. *Social Psychological and Personality Science*, *8*(4), 370–378. https://doi.org/10.1177/1948550617693063

Flemisch, F. O., Bengler, K., Bubb, H., Winner, H., & Bruder, R. (2014). Towards cooperative guidance and control of highly automated vehicles: H-Mode and Conduct-by-Wire. *Ergonomics*, *57*(3), 343–360. https://doi.org/10.1080/00140139.2013.869355

Forster, Y., Naujoks, F., & Neukum, A. (2017). Increasing anthropomorphism and trust in automated driving functions by adding speech output. In *2017 IEEE Intelligent Vehicles Symposium (IV)* (pp. 365–372). IEEE. https://doi.org/10.1109/IVS.2017.7995746

Franz, B., Kauer, M., Geyer, S., & Hakuli, S. (2016). Conduct-by-Wire. In H. Winner, S. Hakuli, F. Lotz, & C. Singer (Eds.), *Handbook of driver assistance systems: Basic information, components and systems for active safety and comfort* (pp. 1483–1497). Cham: Springer Reference.

Fuchs, C., & Diamantopoulos, A. (2009). Using single-item measures for construct measurement in management research: Conceptual issues and application guidelines. *Die Betriebswirtschaft*, *69*(2), 195.

Gasser, T. M. (2012). *Rechtsfolgen zunehmender Fahrzeugautomatisierung: Gemeinsamer Schlussbericht der Projektgruppe* [Legal consequences of an increase in vehicle automation: Collective final report of the project group]. *Berichte der Bundesanstalt für Straßenwesen – Fahrzeugtechnik (F): Vol. 83*. Bremerhaven: Wirtschaftsverlag NW.

Gasser, T. M., & Schmidt, E. A. (2017). *Report on the need for research: Round table on automated driving – Research Working Group*. Retrieved from https://www.bmvi.de/SharedDocs/EN/Documents/VerkehrUndMobilitaet/report-need-for-research.pdf

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, *27*(1), 51–90.

Gelman, A. (2011). Causality and statistical learning. *American Journal of Sociology*, *117*(3), 955–966. https://doi.org/10.1086/662659

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the Technology Acceptance Model to assess automation. *Cognition, Technology & Work*, *14*(1), 39–49. https://doi.org/10.1007/s10111-011-0194-3

Gigerenzer, G. (1998). Surrogates for theories. *Theory & Psychology*, *8*(2), 195–204.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*(5), 587–606. https://doi.org/10.1016/j.socec.2004.09.033

Gigerenzer, G. (2009). Surrogates for theory. *APS Observer*, *22*(2), 21–23.

Gigerenzer, G., & Selten, R. (Eds.). (2002). *Bounded rationality: The adaptive toolbox*. Cambridge, Mass: MIT Press.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Goel, N., Rao, H., Durmer, J., & Dinges, D. (2009). Neurocognitive consequences of sleep deprivation. *Seminars in Neurology*, *29*(04), 320–339. https://doi.org/10.1055/s-0029-1237117

Gold, C. (2016). Modeling of take-over performance in highly automated vehicle guidance (Doctoral dissertation). Technical University of Munich, Munich.

Gold, C., & Bengler, K. (2014). Taking over control from highly automated vehicles. In N. A. Stanton, S. J. Landry, G. Di Bucchianico, & A. Vallicelli (Eds.): *Vol. [8]. Advances in human factors and ergonomics 2014, Advances in human aspects of transportation, Part II* (pp. 64–69). Louisville, KY: AHFE Conference.

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 1938–1942. https://doi.org/10.1177/1541931213571433

Gold, C., Körber, M., Hohenberger, C., Lechner, D., & Bengler, K. (2015). Trust in automation – Before and after the experience of take-over scenarios in a highly automated vehicle. *Procedia Manufacturing*, *3*, 3025–3032. https://doi.org/10.1016/j.promfg.2015.07.847

Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking over control from highly automated vehicles in complex traffic situations: The role of traffic density. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *58*(4), 642–652. https://doi.org/10.1177/0018720816634226

Golias, J., Yannis, G., & Antoniou, C. (2002). Classification of driver-assistance systems according to their impact on road safety and traffic efficiency. *Transport Reviews*, *22*(2), 179–196. https://doi.org/10.1080/01441640110091215

Green, M. (2000). "How long does it take to stop?" Methodological analysis of driver perception-brake times. *Transportation Human Factors*, *2*(3), 195–216. https://doi.org/10.1207/sthf0203_1

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Gründl, M. (2005). Fehler und Fehlverhalten als Ursache von Verkehrsunfällen und Konsequenzen für das Unfallvermeidungspotenzial und die Gestaltung von Fahrerassistenzsystemen [Errors and erratic behavior as a cause for traffic accidents and consequences for the potential of accident avoidance and the design of driver assistance systems] (Doctoral dissertation). University of Regensburg, Regensburg.

Haegerstrom-Portnoy, G., Schneck, M. E., & Brabyn, J. A. (1999). Seeing into old age: Vision function beyond acuity. *Optometry and Vision Science*, *76*(3), 141–158. https://doi.org/10.1097/00006324-199903000-00014

Haines, D. E., & Schenk, M. P. (2015). *Neuroanatomy in clinical context: An atlas of structures, sections, systems, and syndromes* (9th ed.). Philadelphia: Kluwer.

Hakamies-Blomqvist, L. (1998). Older drivers' accident risk: Conceptual and methodological issues. *Accident Analysis & Prevention*, *30*(3), 293–297. https://doi.org/10.1016/S0001-4575(97)00120-6

Hammond, K. R. (1986). Generalization in operational contexts: What does it mean? Can it be done? *IEEE Transactions on Systems, Man, and Cybernetics*, *16*(3), 428–433. https://doi.org/10.1109/TSMC.1986.4308974

Hartley, A. A., & Little, D. M. (1999). Age-related differences and similarities in dual-task interference. *Journal of Experimental Psychology: General*, *128*(4), 416–449.

Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, *16*(3), 319–336. https://doi.org/10.1037/a0024917

Helldin, T., Falkman, G., Riveiro, M., & Davidsson, S. (2013). Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. In J. Terken (Ed.), *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 210–217). New York, NY: ACM. https://doi.org/10.1145/2516540.2516554

Helton, W. S., & Russell, P. N. (2011). Feature absence-presence and two theories of lapses of sustained attention. *Psychological Research*, *75*(5), 384–392. https://doi.org/10.1007/s00426-010-0316-1

Helton, W. S., & Russell, P. N. (2013). Visuospatial and verbal working memory load: Effects on visuospatial vigilance. *Experimental Brain Research*, *224*(3), 429–436. https://doi.org/10.1007/s00221-012-3322-2

Hergeth, S., Lorenz, L., Krems, J. F., & Toenert, L. (2015). Effects of take-over requests and cultural background on automation trust in highly automated driving. In *Proceedings of the 8th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 331–337). Iowa City, Iowa: University of Iowa.

Hergeth, S., Lorenz, L., & Krems, J. F. (2017). Prior familiarization with takeover requests affects drivers' takeover performance and automation trust. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *59*(3), 457–470. https://doi.org/10.1177/0018720816678714

Hergeth, S., Lorenz, L., Vilimek, R., & Krems, J. F. (2016). Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *58*(3), 509–519. https://doi.org/10.1177/0018720815625744

Herslund, M.-B., & Jørgensen, N. O. (2003). Looked-but-failed-to-see-errors in traffic. *Accident Analysis & Prevention*, *35*(6), 885–891. https://doi.org/10.1016/S0001-4575(02)00095-7

Hertzog, C., Kramer, A. F., Wilson, R. S., & Lindenberger, U. (2008). Enrichment effects on adult cognitive development: Can the functional capacity of older adults be preserved and enhanced? *Psychological Science in the Public Interest*, *9*(1), 1–65. https://doi.org/10.1111/j.1539-6053.2009.01034.x

Ho, G., Wheatley, D., & Scialfa, C. T. (2005). Age differences in trust and reliance of a medication management system. *Interacting with Computers*, *17*(6), 690–710. https://doi.org/10.1016/j.intcom.2005.09.007

Hoedemaeker, M., & Brookhuis, K. (1998). Behavioural adaptation to driving with an adaptive cruise control (ACC). *Transportation Research Part F: Traffic Psychology and Behaviour*, *1*(2), 95–106. https://doi.org/10.1016/S1369-8478(98)00008-4

Hoeppner, B. B., Kelly, J. F., Urbanoski, K. A., & Slaymaker, V. (2011). Comparative utility of a single-item versus multiple-item measure of self-efficacy in predicting relapse among young adults. *Journal of Substance Abuse Treatment*, *41*(3), 305–312. https://doi.org/10.1016/j.jsat.2011.04.005

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Hoffman, R. R., Johnson, M., Bradshaw, J. M., & Underbrink, A. (2013). Trust in automation. *IEEE Intelligent Systems*, *28*(1), 84–88. https://doi.org/10.1109/MIS.2013.24

Hogarth, R. M. (2005). The challenge of representative design in psychology and economics. *Journal of Economic Methodology*, *12*(2), 253–263. https://doi.org/10.1080/13501780500086172

Hong, H., & Williamson, A. M. (2008). Speed and accuracy: A complex interplay in skill development. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(8), 682–686. https://doi.org/10.1177/154193120805200816

Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J., & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention*, *38*(1), 185–191. https://doi.org/10.1016/j.aap.2005.09.007

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Horswill, M. S., Marrington, S. A., McCullough, C. M., Wood, J., Pachana, N. A., McWilliam, J., & Raikos, M. K. (2008). The hazard perception ability of older drivers. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *63*(4), 212–218.

Horswill, M. S., & McKenna, F. (2004). Drivers' hazard perception ability: Situation awareness on the road. In S. Banbury & S. Tremblay (Eds.), *A Cognitive approach to situation awareness: Theory and application* (pp. 193–212). Burlington, VT: Ashgate Publishing Ltd.

Howson, C., & Urbach, P. (2006). *Scientific reasoning: The Bayesian approach* (3rd ed.). Chicago: Open Court.

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hubley, A. M., & Zumbo, B. D. (2013). Psychometric characteristics of assessment procedures: An overview. In K. F. Geisinger (Ed.), *APA handbooks in psychology: Vol. 1. Test theory and testing and assessment in industrial and organizational psychology* (pp. 3–20). Washington, D.C.: American Psychological Association.

Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(2), 101–115. https://doi.org/10.1093/geronb/57.2.P101

International Organization for Standardization (2012). *Road vehicles – Ergonomic aspects of transport information and control systems – Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems.* (ISO Standard No. 14198): BSI.

Irmscher, M., & Ehmann, M. (2004). *Driver classification using ve-DYNA advanced driver* (Technical Report No. 2004-01-0451).

Ivers, R., Senserrick, T., Boufous, S., Stevenson, M. R., Chen, H.-Y., Woodward, M., & Norton, R. (2009). Novice drivers' risky driving behavior, risk perception, and crash risk: Findings from the DRIVE study. *American Journal of Public Health*, *99*(9), 1638–1644. https://doi.org/10.2105/AJPH.2008.150367

Jagacinski, R. J., & Flach, J. (2003). *Control theory for humans: Quantitative approaches to modeling performance.* Mahwah, N.J: L. Erlbaum Associates.

JASP Team. (2018). JASP (Version 0.8.1) [Computer software]. Retrieved from https://jasp-stats.org/

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

Kahneman, D. (1973). *Attention and effort. Prentice-Hall series in experimental psychology*. Englewood Cliffs, N.J: Prentice Hall.

Kahneman, D., Ben-Ishai, R., & Lotan, M. (1973). Relation of a test of attention to road accidents. *Journal of Applied Psychology*, *58*(1), 113–115. https://doi.org/10.1037/h0035426

Kämpchen, N., Aeberhard, M., Ardelt, M., & Rauch, S. (2012). Technologies for highly automated driving on highways. *ATZ Worldwide*, *114*(6), 34–38. https://doi.org/10.1007/s38311-012-0176-y

Karlen, W., Cardin, S., Thalmann, D., & Floreano, D. (2010). Enhancing pilot performance with a SymBodic system. In *Annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 6599–6602). Piscataway, NJ: IEEE. https://doi.org/10.1109/IEMBS.2010.5627127

Kelava, A., & Schermelleh-Engel, K. (2012). Latent-State-Trait-Theorie (LST-Theorie) [Latent state-trait theory]. In H. Moosbrugger & A. Kelava (Eds.), *Springer-Lehrbuch. Testtheorie und Fragebogenkonstruktion* (2nd ed., pp. 344–360). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

Keller, D., & Rice, S. (2009). System-wide versus component-specific trust using multiple aids. *The Journal of General Psychology*, *137*(1), 114–128. https://doi.org/10.1080/00221300903266713

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2014). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, *44*(3), 486–507. https://doi.org/10.1177/0049124114543236

Kienle, M., Damböck, D., Kelsch, J., Flemisch, F. O., & Bengler, K. (2009). Towards an H-Mode for highly automated vehicles. In A. Schmidt, A. Dey, T. Seder, & O. Juhlin (Eds.), *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '09* (pp. 19–23). New York, New York, USA: ACM Press. https://doi.org/10.1145/1620509.1620513

Kimball, A. W. (1957). Errors of the third kind in statistical consulting. *Journal of the American Statistical Association*, *52*(278), 133–142. https://doi.org/10.2307/2280840

Kircher, K. (2007). *Driver distraction – A review of the literature* (VTI rapport No. 594A). Linköping, Sweden.

Kirlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why an “aid” can (and should) go unused. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *35*(2), 221–242. https://doi.org/10.1177/001872089303500203

Kirschbaum, M. (2015). Highly automated driving for commercial vehicles. In P. Pfeffer (Ed.), *Proceedings. 6th International Munich Chassis Symposium 2015* (pp. 5–15). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-09711-0_2

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data* (Technical Report No. DOT HS 810 594). Washington, USA.

König, C. J., Bühner, M., & Mürling, G. (2005). Working memory, fluid intelligence, and attention are predictors of multitasking performance, but polychronicity and extraversion are not. *Human Performance*, *18*(3), 243–266. https://doi.org/10.1207/s15327043hup1803_3

Körber, M. (2019). Theoretical considerations and development of a questionnaire to measure trust in automation. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018): Volume VI:*

Transport Ergonomics and Human Factors (TEHF), Aerospace Human Factors and Ergonomics (1st ed., pp. 13–30). Springer.

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, *66*, 18–31. https://doi.org/10.1016/j.apergo.2017.07.006

Körber, M., & Bengler, K. (2014). Potential individual differences regarding automation effects in automated driving. In C. S. G. González, C. C. Ordóñez, & H. Fardoun (Eds.), *Interacción 2014: Proceedings of the XV International Conference on Human Computer Interaction* (pp. 152–158). New York, NY, USA: ACM. https://doi.org/10.1145/2662253.2662275

Körber, M., Cingel, A., Zimmermann, M., & Bengler, K. (2015). Vigilance decrement and passive fatigue caused by monotony in automated driving. *Procedia Manufacturing*, *3*, 2403–2409. https://doi.org/10.1016/j.promfg.2015.07.499

Körber, M., Gold, C., Lechner, D., & Bengler, K. (2016). The influence of age on the take-over of vehicle control in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *39*, 19–32. https://doi.org/10.1016/j.trf.2016.03.002

Körber, M., Prasch, L., & Bengler, K. (2018). Why do I have to drive now? Post hoc explanations of takeover requests. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *60*(3), 305-323. https://doi.org/10.1177/0018720817747730

Körber, M., Radlmayr, J., & Bengler, K. (2016). Bayesian highest density intervals of take-over times for highly automated driving in different traffic densities. In *Proceedings of the Human Factors and Ergonomics Society (HFES) 60th Annual Meeting 2016* (Vol. 60, pp. 2009–2013). https://doi.org/10.1177/1541931213601457

Körber, M., Schneider, W., & Zimmermann, M. (2015). Vigilance, boredom proneness and detection time of a malfunction in partially automated driving. In *International Conference on Collaboration Technologies and Systems (CTS)* (pp. 70–76). IEEE. https://doi.org/10.1109/CTS.2015.7210402

Körber, M., Weißgerber, T., Kalb, L., Blaschke, C., & Farid, M. (2015). Prediction of take-over time in highly automated driving by two psychometric tests. *Dyna*, *82*(193), 195–201. https://doi.org/10.15446/dyna.v82n193.53496

Koustanaï, A., Boloix, E., van Elslande, P., & Bastien, C. (2008). Formation of expectations while driving: Influence of the possibility and the necessity to anticipate on the ability to identify danger. *Transportation Research Part F: Traffic Psychology and Behaviour*, *11*(2), 147–157. https://doi.org/10.1016/j.trf.2007.09.001

Kramer, A. F., Hahn, S., & Gopher, D. (1999). Task coordination and aging: Explorations of executive control processes in the task switching paradigm. *Acta Psychologica*, *101*(2-3), 339–378.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (2007). *Additive and polynomial representations. Foundations of measurement: Vol. 1*. Mineola, NY: Dover Publ.

Kray, J., Eber, J., & Lindenberger, U. (2004). Age differences in executive functioning across the lifespan: The role of verbalization in task preparation. *Acta Psychologica*, *115*(2-3), 143–165. https://doi.org/10.1016/j.actpsy.2003.12.001

Kray, J., Li, K. Z.H., & Lindenberger, U. (2002). Age-related changes in task-switching components: The role of task uncertainty. *Brain and Cognition*, *49*(3), 363–381. https://doi.org/10.1006/brcg.2001.1505

Kray, J., & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and Aging*, *15*(1), 126–147. https://doi.org/10.1037/0882-7974.15.1.126

Krings, T., Töpper, R., Foltys, H., Erberich, S., Sparing, R., Willmes, K., & Thron, A. (2000). Cortical activation patterns during complex motor tasks in piano players and control subjects. A functional magnetic resonance imaging study. *Neuroscience Letters*, *278*(3), 189–193. https://doi.org/10.1016/S0304-3940(99)00930-1

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). San Diego, CA: Academic Press.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

Van der Laan, J. D., Heino, A., & de Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, *5*(1), 1–10. https://doi.org/10.1016/S0968-090X(96)00025-3

Lachman, R., Lachman, J. L., & Butterfield, E. C. (1979). *Cognitive psychology and information processing: An introduction*. Mahwah: Taylor and Francis. Retrieved from http://gbv.eblib.com/patron/FullRecord.aspx?p=4219171

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Larson, G. E., & Merritt, C. R. (1991). Can accidents be predicted? An empirical test of the cognitive failures questionnaire. *Applied Psychology*, *40*(1), 37–45. https://doi.org/10.1111/j.1464-0597.1991.tb01356.x

Larue, G. S., Rakotonirainy, A., & Pettitt, A. N. (2011). Driving performance impairments due to hypovigilance on monotonous roads. *Accident Analysis & Prevention*, *43*(6), 2037–2046. https://doi.org/10.1016/j.aap.2011.05.023

Lee, J. D. (2008). Review of a pivotal human factors article: "Humans and automation: Use, misuse, disuse, abuse". *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *50*(3), 404–410. https://doi.org/10.1518/001872008x288547

Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243–1270. https://doi.org/10.1080/00140139208967392

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, *40*(1), 153–184. https://doi.org/10.1006/ijhc.1994.1007

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Levin, S., & Woolf, N. (2017, July 1). Tesla driver killed while using autopilot was watching Harry Potter, witness says. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2016/jul/01/tesla-driver-killed-autopilot-self-driving-car-harry-potter

Levinson, J., & Thrun, S. (2010). Robust vehicle localization in urban environments using probabilistic maps. In *2010 IEEE International Conference on Robotics and Automation* (pp. 4372–4378). IEEE. https://doi.org/10.1109/ROBOT.2010.5509700

Lewandowsky, S., Mundy, M., & Tan, G. P. A. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied*, *6*(2), 104–123. https://doi.org/10.1037/1076-898X.6.2.104

Lobb, M. L., & Stern, J. A. (2009). Pattern of eyelid motion predictive of decision errors during drowsiness: Oculomotor indices of altered states. *International Journal of Neuroscience*, *30*(1-2), 17–22. https://doi.org/10.3109/00207458608985650

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18*(5), 883–914.

Luczak, H. (1975). *Untersuchungen informatorischer Belastung und Beanspruchung des Menschen* [Investigation of informational stress and strain on humans]. *Fortschrittsberichte der der VDI-Zeitschriften: 10(2).* Düsseldorf: VDI-Verlag.

Luczak, H. (1998). *Arbeitswissenschaft* [Ergonomics]. Berlin, Heidelberg, s.l.: Springer. Retrieved from http://dx.doi.org/10.1007/978-3-662-05831-2

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84

Mackenzie, A. K., & Harris, J. M. (2015). Eye movements and hazard perception in active and passive driving. *Visual Cognition*, *23*(6), 736–757. https://doi.org/10.1080/13506285.2015.1079583

Mackenzie, A. K., & Harris, J. M. (2017). A link between attentional function, effective eye movements, and driving ability. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(2), 381–394. https://doi.org/10.1037/xhp0000297

Madhavan, P., & Wiegmann, D. A. (2005). Cognitive anchoring on self-generated decisions reduces operator reliance on automated diagnostic aids. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *47*(2), 332–341. https://doi.org/10.1518/0018720054679489

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301. https://doi.org/10.1080/14639220500337708

Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. In *Proceedings of the 11th Australasian Conference on Information Systems* (pp. 6–8).

Maehigashi, A., Miwa, K., Terai, H., Kojima, K., & Morita, J. (2012). Experimental investigation of relationship between complacency and tendency to use automation system. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1960–1965).

Mahr, A., & Müller, C. (2011). A schema of possible negative effects of advanced driver assistant systems. In *Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 116–121). Iowa City, Iowa: The University of Iowa, Public Policy Center.

Manzey, D. (2012). Systemgestaltung und Automatisierung [System design and automation]. In P. Badke-Schaub, G. Hofinger, & K. Lauche (Eds.), *Human Factors* (pp. 333–352). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-19886-1_19

Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of*

Cognitive Engineering and Decision Making, *6*(1), 57–87. https://doi.org/10.1177/1555343411433844

Marberger, C., Mielenz, H., Naujoks, F., Radlmayr, J., Bengler, K., & Wandtner, B. (2018). Understanding and applying the concept of "driver availability" in automated driving. In N. A. Stanton (Ed.), *Advances in Intelligent Systems and Computing: Vol. 597. Advances in human aspects of transportation: Proceedings of the AHFE 2017 International Conference on Human Factors in Transportation, July 17−21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA* (pp. 595–605). Cham: Springer.

Markkula, G., Benderius, O., Wolff, K., & Wahde, M. (2012). A review of near-collision driver behavior models. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *54*(6), 1117–1143. https://doi.org/10.1177/0018720812448474

Masunaga, H., & Horn, J. (2001). Expertise and age-related changes in components of intelligence. *Psychology and Aging*, *16*(2), 293–311. https://doi.org/10.1037//0882-7974.16.2.293

Mathias, J. L., & Lucas, L. K. (2009). Cognitive predictors of unsafe driving in older drivers: A meta-analysis. *International Psychogeriatrics*, *21*(4), 637–653. https://doi.org/10.1017/S1041610209009119

Matthews, G., & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *55*(2), 659–686. https://doi.org/10.1080/02724980143000505

Matthews, G., Joyner, L. A., Gilliland, K., Huggins, J., & Falconer, S. (1999). Validation of a comprehensive stress state questionnaire: Towards a state big three? In I. Mervielde, I. J. Deary, F. de Fruyt, & I. Ostendorf (Eds.), *Personality psychology in Europe* (pp. 335–350). Tilburg: Tilburg University Press.

Matthews, G., Sparkes, T. J., & Bygrave, H. M. (1996). Attentional overload, stress, and simulated driving performance. *Human Performance*, *9*(1), 77–101. https://doi.org/10.1207/s15327043hup0901_5

Matthews, G., Warm, J. S., Shaw, T. H., & Finomore, V. S. (2010). A multivariate test battery for predicting vigilance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(14), 1072–1076. https://doi.org/10.1177/154193121005401405

Maurer, M., Gerdes, J. C., Lenz, B., & Winner, H. (2015). *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte* [Autonomous driving: Technical, legal, and societal aspects]. Berlin: Springer Vieweg.

May, J. F., & Baldwin, C. L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*(3), 218–224. https://doi.org/10.1016/j.trf.2008.11.005

Maycock, G., Lockwood, C. M., & Lester, J. F. (1991). *The accident liability of car drivers* (No. 315). Crowthorne.

Mayer, A. K., Rogers, W. A., & Fisk, A. D. (2009). *Understanding technology acceptance: Effects of user expectancies on human-automation interaction* (Technical Report HFA-TR-09-07). Atlanta, GA.

Mayer, A. K., Sanchez, J., Fisk, A. D., & Rogers, W. A. (2006). Don't let me down: The role of operator expectations in human-automation interaction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(21), 2345–2349. https://doi.org/10.1177/154193120605002110

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.5465/AMR.1995.9508080335

Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, *84*(1), 123–136. https://doi.org/10.1037//0021-9010.84.1.123

Mayr, U. (2001). Age differences in the selection of mental sets: The role of inhibition, stimulus ambiguity, and response-set overlap. *Psychology and Aging*, *16*(1), 96–109. https://doi.org/10.1037/0882-7974.16.1.96

McCartt, A. T., Mayhew, D. R., Braitman, K. A., Ferguson, S. A., & Simpson, H. M. (2009). Effects of age and experience on young driver crashes: Review of recent literature. *Traffic Injury Prevention*, *10*(3), 209–219. https://doi.org/10.1080/15389580802677807

McCoach, D. B. (2003). SEM isn't just the Schoolwide Enrichment Model anymore: Structural equation modeling (SEM) in gifted education. *Journal for the Education of the Gifted*, *27*(1), 36–61. https://doi.org/10.1177/016235320302700104

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain: School and corporate applications* (3rd ed.). New York, NY: Springer. Retrieved from http://dx.doi.org/10.1007/978-1-4614-7135-6

McGuirl, J. M., & Sarter, N. B. (2006). Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *48*(4), 656–665. https://doi.org/10.1518/001872006779166334

McKnight, A. J., & McKnight, A. S. (1999). Multivariate analysis of age-related driver ability and performance deficits. *Accident Analysis & Prevention*, *31*(5), 445–454. https://doi.org/10.1016/S0001-4575(98)00082-7

McKnight, D. H., & Chervany, N. L. (1996). *The meanings of trust* (WP No. 96-04).

McKnight, D. H., & Chervany, N. L. (2001). Trust and distrust definitions: One bite at a time. In R. Falcone, M. Singh, & Y.-H. Tan (Eds.), *Lecture notes in computer science: Vol. 2246. Trust in Cyber-societies: Integrating the Human and Artificial Perspectives* (pp. 27–54). Berlin, Heidelberg: Springer.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

McPherson, S. L. (1999). Expert-novice differences in performance skills and problem representations of youth and adults during tennis competition. *Research Quarterly for Exercise and Sport*, *70*(3), 233–251. https://doi.org/10.1080/02701367.1999.10608043

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*(2), 103–115. https://doi.org/10.1086/288135

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *50*(2), 194–210. https://doi.org/10.1518/001872008X288574

Merritt, S. M., Unnerstall, J. L., Lee, D., & Huber, K. (2015). Measuring individual differences in the perfect automation schema. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *57*(5), 740–753. https://doi.org/10.1177/0018720815581247

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *43*(4), 563–572. https://doi.org/10.1518/001872001775870395

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *46*(2), 196–204.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, *88*(3), 355–383. https://doi.org/10.1111/j.2044-8295.1997.tb02641.x

Middleton, H., Westwood, D., Robson, J., & Kok, D. (2005). Assessment and decision criteria for driving competence in the elderly. In G. Underwood (Ed.), *Traffic and transport psychology: Theory and application* (pp. 101–113). Amsterdam: Elsevier.

Mihal, W. L., & Barrett, G. V. (1976). Individual differences in perceptual information processing and their relation to automobile accident involvement. *Journal of Applied Psychology*, *61*(2), 229–233. https://doi.org/10.1037/0021-9010.61.2.229

Miller, S. M., Taylor-Piliae, R. E., & Insel, K. C. (2016). The association of physical activity, cognitive processes and automobile driving ability in older adults: A review of the literature. *Geriatric Nursing (New York, N.Y.)*, *37*(4), 313–320. https://doi.org/10.1016/j.gerinurse.2016.05.004

Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *38*(2), 311–322. https://doi.org/10.1177/001872089606380211

Molnar, L. J., & Eby, D. W. (2008). The relationship between self-regulation and driving-related abilities in older drivers: An exploratory study. *Traffic Injury Prevention*, *9*(4), 314–319. https://doi.org/10.1080/15389580801895319

Monk, C. A., Boehm-Davis, D. A., & Trafton, J. G. (2004). Recovering from interruptions: Implications for driver distraction research. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *46*(4), 650–663.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7

Moosbrugger, H., & Kelava, A. (Eds.). (2012). *Testtheorie und Fragebogenkonstruktion* [Test theory and construction of questionnaires] (2nd ed.). *Springer-Lehrbuch*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

Morgan, B., D'Mello, S., Adams, M. J., Radvansky, G., Haass, M., & Tamplin, A. (2013). Individual differences in multitasking ability and adaptability. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *55*(4), 776–788. https://doi.org/10.1177/0018720812470842

Morgan, B., D'Mello, S., Abbott, R., Haass, M., Tamplin, A., & Radvansky, G. (2013). Performance-based adaptability profiles in multitasking. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 843–847. https://doi.org/10.1177/1541931213571183

Morse, C. K. (1993). Does variability increase with age? An archival study of cognitive measures. *Psychology and Aging*, *8*(2), 156–164. https://doi.org/10.1037/0882-7974.8.2.156

Mosier, K. L., & Skitka, L. J. (1996). Human decision-makers and automated decision aids: Made for each other? In R. Parasuraman & M. Mouloua (Eds.), *Human factors in transportation. Automation and human performance: Theory and applications* (pp. 201–220). Mahwah, NJ: Erlbaum.

Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL test Battery. *Journal of Neuroscience Methods*, *222*, 250–259. https://doi.org/10.1016/j.jneumeth.2013.10.024

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, *37*(11), 1905–1922. https://doi.org/10.1080/00140139408964957

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*(3), 429–460. https://doi.org/10.1080/00140139608964474

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, *27*(5-6), 527–539. https://doi.org/10.1016/S0020-7373(87)80013-5

Mulder, M., Abbink, D. A., & Boer, E. R. (2012). Sharing control with haptics: Seamless driver support from manual to automatic control. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *54*(5), 786–798. https://doi.org/10.1177/0018720812443984

Mullen, N., Charlton, J., Devlin, A., & Bédard, M. (2011). Simulator validity: Behaviors observed on the simulator and on the road. In D. L. Fisher, M. Rizzo, J. Caird, & J. D. Lee (Eds.), *Handbook of driving simulation for engineering, medicine, and psychology* (13-1–13-18). Boca Raton: CRC Press.

Nagy, M. S. (2002). Using a single-item approach to measure facet job satisfaction. *Journal of Occupational and Organizational Psychology*, *75*(1), 77–86. https://doi.org/10.1348/096317902167658

Najm, W. G., Stearns, M. D., Howarth, H., Koopmann, J., & Hitz, J. (2006). *Evaluation of an automotive rear-end collision avoidance system* (No. DOT HS 810 569). Washington, DC.

National Highway Traffic Safety Administration. (2013). *Preliminary statement of policy concerning automated vehicles*. Washington, DC.

Naujoks, F., Purucker, C., Neukum, A., Wolter, S., & Steiger, R. (2015). Controllability of partially automated driving functions – Does it matter whether drivers are allowed to take their hands off the steering wheel? *Transportation Research Part F: Traffic Psychology and Behaviour*, *35*, 185–198. https://doi.org/10.1016/j.trf.2015.10.022

Naujoks, F., & Totzke, I. (2014). Behavioral adaptation caused by predictive warning systems – The case of congestion tail warnings. *Transportation Research Part F: Traffic Psychology and Behaviour*, *26*, 49–61. https://doi.org/10.1016/j.trf.2014.06.010

Neumann, O. (1996). Theorien der Aufmerksamkeit [Theories of attention]. In O. Neumann & A. F. Sanders (Eds.), *Enzyklopädie der Psychologie: Themenbereich C Theorie und Forschung - Kognition: Vol. 2. Aufmerksamkeit* (pp. 559–644). Göttingen: Hogrefe Verlag für Psychologie.

Norman, D. A. (1990). The 'problem' with automation: Inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *327*(1241), 585–593. https://doi.org/10.1098/rstb.1990.0101

Norman, D. A. (2003). *Things that make us smart: Defending human attributes in the age of the machine*. Reading, Mass.: Perseus Books.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). *McGraw-Hill*. New York, NY: McGraw-Hill.

O'Brien, F., Klauer, S. G., Ehsani, J., & Simons-Morton, B. G. (2016). Changes over 12 months in eye glances during secondary task engagement among novice drivers. *Accident Analysis & Prevention*, *93*, 48–54. https://doi.org/10.1016/j.aap.2016.04.022

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory. *Intelligence*, *31*(2), 167–193. https://doi.org/10.1016/S0160-2896(02)00115-0

Office of Defects Investigation. (2017). *ODI resume investigation PE 16-007* (PE No. 16-007).

Oliver, B. M., & Billingham, J. (1971). Project Cyclops: A design study of a system for detecting extraterrestrial intelligent life. In *The 1971 NASA/ASEE Summer Fac. Fellowship Program (NASA-CR-114445)* (Vol. 1).

Oltedal, S., & Rundmo, T. (2006). The effects of personality and gender on risky driving behaviour and accident involvement. *Safety Science*, *44*(7), 621–628. https://doi.org/10.1016/j.ssci.2005.12.003

Onnasch, L. (2015). Benefits and costs of automation support: The role of function allocation and automation reliability (Doctoral dissertation). Technische Universität Berlin, Berlin. Retrieved from http://dx.doi.org/10.14279/depositonce-4407

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *56*(3), 476–488. https://doi.org/10.1177/0018720813501549

Onnasch, L., Wiczorek, R., & Manzey, D. (2011). Ist „Wenig-Vertrauen" gleich Misstrauen? Zur (UN-)Abhängigkeit von Vertrauen und Misstrauen in Alarmsysteme [Is "low trust" the same as mistrust? About the (in)depence of trust and mistrust in alarm systems]. In S. Schmid, M. Elepfandt, J. Adenauer, & A., Lichtenstein (Eds.): *Vol. 33. Berichte aus dem Zentrum Mensch-Maschine-Systeme der Technischen Universität Berlin, Reflexionen und Visionen der Mensch-Maschine-Interaktion - aus der Vergangenheit lernen, Zukunft gestalten: 9. Berliner Werkstatt Mensch-Maschine-Systeme* (pp. 615–620). Düsseldorf: VDI-Verlag.

Parasuraman, R. (2000). Designing automation for human use: Empirical studies and quantitative models. *Ergonomics*, *43*(7), 931–951. https://doi.org/10.1080/001401300409125

Parasuraman, R. (2009). Assaying individual differences in cognition with molecular genetics: Theory and application. *Theoretical Issues in Ergonomics Science*, *10*(5), 399–416. https://doi.org/10.1080/14639220903106403

Parasuraman, R. (2011). Neuroergonomics: Brain, cognition, and performance at work. *Current Directions in Psychological Science*, *20*(3), 181–186. https://doi.org/10.1177/0963721411409176

Parasuraman, R., & Jiang, Y. (2012). Individual differences in cognition, affect, and performance: Behavioral, neuroimaging, and molecular genetic approaches. *NeuroImage*, *59*(1), 70–82. https://doi.org/10.1016/j.neuroimage.2011.04.040

Parasuraman, R., & Manzey, D. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced 'complacency'. *The International Journal of Aviation Psychology*, *3*(1), 1–23. https://doi.org/10.1207/s15327108ijap0301_1

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Parasuraman, R., & Rizzo, M. (Eds.). (2008). *Neuroergonomics: The brain at work. Oxford series in human-technology interaction*. Oxford: Oxford University Press.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, *2*(2), 140–160. https://doi.org/10.1518/155534308X284417

Parasuraman, R., de Visser, E., Lin, M.-K., & Greenwood, P. M. (2012). Dopamine beta hydroxylase genotype identifies individuals less susceptible to bias in computer-assisted decision making. *PLoS ONE*, *7*(6), e39675. https://doi.org/10.1371/journal.pone.0039675

Park, D. C., & Gutchess, A. H. (2012). Cognitive aging and everyday life. In D. C. Park & N. Schwarz (Eds.), *Cognitive aging: A primer* (pp. 217–232). Hoboken: Taylor and Francis.

Pashler, H., Yantis, S., Medin, D., Gallistel, R., & Wixted, J. T. (Eds.). (2004). *Stevens' handbook of experimental psychology* (3rd ed.): John Wiley & Sons.

Pashler, H., & Baylis, G. C. (1991). Procedural learning: I. Locus of practice effects in speeded choice tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 20–32. https://doi.org/10.1037/0278-7393.17.1.20

Pattyn, N., Neyt, X., Henderickx, D., & Soetens, E. (2008). Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue? *Physiology & Behavior*, *93*(1-2), 369–378. https://doi.org/10.1016/j.physbeh.2007.09.016

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). New York, NY: The Guilford Press.

Paulweber, M. (2017). Validation of highly automated safe and secure systems. In D. Watzenig & M. Horn (Eds.), *Automated driving: Safer and more efficient future driving* (pp. 437–450). Cham, s.l.: Springer International Publishing.

Payre, W., Cestac, J., Dang, N.-T., Vienne, F., & Delhomme, P. (2017). Impact of training and in-vehicle task performance on manual control recovery in an automated car. *Transportation Research Part F: Traffic Psychology and Behaviour*, *46*, 216–227. https://doi.org/10.1016/j.trf.2017.02.001

Payre, W., Cestac, J., & Delhomme, P. (2014). Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 252–263. https://doi.org/10.1016/j.trf.2014.04.009

Payre, W., Cestac, J., & Delhomme, P. (2016). Fully automated driving: Impact of trust and practice on manual control recovery. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *58*(2), 229–241. https://doi.org/10.1177/0018720815612319

Petermann-Stock, I., Hackenberg, L., Muhr, T., & Mergl, C. (2013). Wie lange braucht der Fahrer? Eine Analyse zu Übernahmezeiten aus verschiedenen Nebentätigkeiten während einer hochautomatisierten Staufahrt [How much time does the driver need? An analysis of take-over times at different secondary tasks during a highly automated drive in a traffic jam.]. In *6. Tagung Fahrerassistenz: Der Weg zum automatischen Fahren.* TÜV SÜD Akademie GmbH.

Petermeijer, S. M., Abbink, D. A., & de Winter, J. C. F. (2015). Should drivers be operating within an automation-free bandwidth? Evaluating haptic steering support systems with different levels of authority. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *57*(1), 5–20. https://doi.org/10.1177/0018720814563602

Petermeijer, S. M., Cieler, S., & de Winter, J. C. F. (2017). Comparing spatially static and dynamic vibrotactile take-over requests in the driver seat. *Accident Analysis & Prevention*, *99*(Pt A), 218–227. https://doi.org/10.1016/j.aap.2016.12.001

Petermeijer, S. M., de Winter, J. C. F., & Bengler, K. (2016). Vibrotactile displays: A survey with a view on highly automated driving. *IEEE Transactions on Intelligent Transportation Systems*, *17*(4), 897–907. https://doi.org/10.1109/TITS.2015.2494873

Platt, J. R. (1964). Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science*, *146*(3642), 347–353. https://doi.org/10.1126/science.146.3642.347

Pöhler, G., Heine, T., & Deml, B. (2016). Itemanalyse und Faktorstruktur eines Fragebogens zur Messung von Vertrauen im Umgang mit automatischen Systemen [Analysis of items and factor structure of a questionnaire for the measurement of trust in the interaction with automated systems]. *Zeitschrift Für Arbeitswissenschaft*, *70*(3), 151–160. https://doi.org/10.1007/s41449-016-0024-9

Ponds, R. W. H. M., Brouwer, W. H., & Van Wolffelaar, P. C. (1988). Age differences in divided attention in a simulated driving task. *Journal of Gerontology*, *43*(6), 151–156. https://doi.org/10.1093/geronj/43.6.P151

Pop, V. L., Shrewsbury, A., & Durso, F. T. (2015). Individual differences in the calibration of trust in automation. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *57*(4), 545–556. https://doi.org/10.1177/0018720814564422

Popper, K. R. (1959). *The logic of scientific discovery*. Oxford, England: Basic Books.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology*, *109*(2), 160–174.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Poulton, E. C. (1950). Perceptual anticipation and reaction time. *Quarterly Journal of Experimental Psychology*, *2*(3), 99–112. https://doi.org/10.1080/17470215008416582

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, *48*(1), 28–56. https://doi.org/10.1080/00273171.2012.710386

Pretz, J. E. (2008). Intuition versus analysis: Strategy and experience in complex everyday problem solving. *Memory & Cognition*, *36*(3), 554–566. https://doi.org/10.3758/MC.36.3.554

Pretz, J. E., & Sternberg, R. J. (Eds.). (2005). *Cognition and intelligence: Identifying the mechanisms of the mind*. Cambridge, U.K, New York: Cambridge University Press.

Prinzel, L. J. I., DeVries, H., Freeman, F. G., & Mikulka, P. J. (2001). *Examination of automation-induced complacency and individual difference variates*. Hampton, VA: NASA Langley Research Center.

Proctor, R. W., & van Zandt, T. (2008). *Human factors in simple and complex systems* (2nd ed.). Hoboken: CRC Press.

Proctor, R. W., & Vu, K.-P. L. (2009). Human information processing: An overview for human–computer interaction. In A. Sears & J. A. Jacko (Eds.), *Human factors and ergonomics. Human-computer interaction: Fundamentals* (pp. 19–38). Boca Raton: CRC Press.

Punke, M., Menzel, S., Werthessen, B., Stache, N., & Höpfl, M. (2016). Automotive camera (hardware). In H. Winner, S. Hakuli, F. Lotz, & C. Singer (Eds.), *Handbook of driver assistance systems: Basic information, components and systems for active safety and comfort* (pp. 431–460). Cham: Springer Reference.

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 2063–2067. https://doi.org/10.1177/1541931214581434

Radlmayr, J., Körber, M., Feldhütter, A., & Bengler, K. (2016). Methoden und Fahrermodelle für Hochautomatisiertes Fahren [Methods and driver models for highly automated driving]. In K. Kompaß (Ed.), *Haus der Technik Fachbuch: Band 144. Methodenentwicklung für aktive Sicherheit und automatisiertes Fahren: 2. Expertendialog zu Wirksamkeit - Beherrschbarkeit - Absicherung.* Renningen: expert Verlag.

Rajaonah, B., Tricot, N., Anceaux, F., & Millot, P. (2008). The role of intervening variables in driver–ACC cooperation. *International Journal of Human-Computer Studies*, *66*(3), 185–197. https://doi.org/10.1016/j.ijhcs.2007.09.002

Rao, H., Korczykowski, M., Pluta, J., Hoang, A., & Detre, J. A. (2008). Neural correlates of voluntary and involuntary risk taking in the human brain: An fMRI Study of the Balloon Analog Risk Task (BART). *NeuroImage*, *42*(2), 902–910. https://doi.org/10.1016/j.neuroimage.2008.05.046

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-13*(3), 257–266. https://doi.org/10.1109/TSMC.1983.6313160

Rauch, S., Aeberhard, M., Ardelt, M., & Kämpchen, N. (2012). Autonomes Fahren auf der Autobahn – eine Potentialstudie für zukünftige Fahrerassistenzsysteme [Autonomous driving on highways – a study on the potential of future driver assistance systems]. In *5. Tagung Fahrerassistenz: Schwerpunkt Vernetzung.* TÜV SÜD Akademie GmbH.

Reason, J. (2000). Human error: Models and management. *BMJ (Clinical Research Ed.)*, *320*(7237), 768–770. https://doi.org/10.1136/bmj.320.7237.768

Recarte, M. A., & Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied*, *6*(1), 31–43.

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied*, *9*(2), 119–137. https://doi.org/10.1037/1076-898X.9.2.119

Regan, M. A., Horberry, T., & Stevens, A. (Eds.). (2014). *Driver acceptance of new technology: Theory, measurement and optimisation. Human factors in road and rail transport.* Farnham, Surrey, England, UK: Ashgate Publishing Ltd.

Reichenbach, J., Onnasch, L., & Manzey, D. (2010). Misuse of automation: The impact of system experience on complacency and automation bias in interaction with automated aids. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(4), 374–378. https://doi.org/10.1177/154193121005400422

Reimer, B. (2014). Driver assistance systems and the transition to automated vehicles: A path to increase older adult safety and mobility? *Public Policy & Aging Report*, *24*(1), 27–31. https://doi.org/10.1093/ppar/prt006

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, *49*(1), 95–112. https://doi.org/10.1037/0022-3514.49.1.95

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Riley, V. (1995). What avionics engineers should know about pilots and automation. In *Proceedings of 14th Digital Avionics Systems Conference* (pp. 252–257). IEEE. https://doi.org/10.1109/dasc.1995.482836

Roberts, J., Hodgson, R., & Dolan, P. (2011). "It's driving her mad": Gender differences in the effects of commuting on psychological health. *Journal of Health Economics*, *30*(5), 1064–1076. https://doi.org/10.1016/j.jhealeco.2011.07.006

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358–367.

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, *27*(2), 151–161. https://doi.org/10.1177/0146167201272002

Robinson, M. D., & Tamir, M. (2005). Neuroticism as mental noise: A relation between neuroticism and reaction time standard deviations. *Journal of Personality and Social Psychology*, *89*(1), 107–114. https://doi.org/10.1037/0022-3514.89.1.107

Robson, K., Plangger, K., Kietzmann, J. H., McCarthy, I., & Pitt, L. (2015). Is it all a game? Understanding the principles of gamification. *Business Horizons*, *58*(4), 411–420. https://doi.org/10.1016/j.bushor.2015.03.006

Rohmert, W. (1984). Das Belastungs-Beanspruchungs-Konzept [The concept of stress and strain]. *Zeitschrift Für Arbeitswissenschaft*, *38*(4), 193–200.

Rosenbaum, D. A. (1980). Human movement initiation: Specification of arm, direction, and extent. *Journal of Experimental Psychology: General*, *109*(4), 444–474.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*(4), 651–665. https://doi.org/10.1111/j.1467-6494.1967.tb01454.x

Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist*, *26*(5), 443–452. https://doi.org/10.1037/h0031464

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rovira, E., Pak, R., & McLaughlin, A. C. (2016). Effects of individual differences in working memory on performance and trust with various degrees of automation. *Theoretical Issues in Ergonomics Science*, *73*, 1–19. https://doi.org/10.1080/1463922X.2016.1252806

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

Rupp, M. A., Gentzler, M. D., & Smither, J. A. (2016). Driving under the influence of distraction: Examining dissociations between risk perception and engagement in distracted driving. *Accident Analysis & Prevention*, *97*, 220–230. https://doi.org/10.1016/j.aap.2016.09.003

SAE International (2016, September 30). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicle.* (Standard, J3016_201609).

Sakaluk, J. K., & Short, S. D. (2017). A methodological review of exploratory factor analysis in sexuality research: Used practices, best practices, and data analysis resources. *Journal of Sex Research*, *54*(1), 1–9. https://doi.org/10.1080/00224499.2015.1137538

Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science*, *2*(3), 179–183. https://doi.org/10.1111/j.1467-9280.1991.tb00127.x

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*(4), 507–514. https://doi.org/10.1016/j.neurobiolaging.2008.09.023

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), 763–776. https://doi.org/10.1037/0012-1649.27.5.763

Salvucci, D. D. (2013). Multitasking. In J. D. Lee & M. J. Dainoff (Eds.), *Oxford library of psychology. The Oxford handbook of cognitive engineering* (pp. 57–67). New York, NY: Oxford University Press.

Sanchez, J., Fisk, A. D., & Rogers, W. A. (2004). Reliability and age-related effects on trust and reliance of a decision support aid. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(3), 586–589. https://doi.org/10.1177/154193120404800366

Sanders, A. F. (1983). Towards a model of stress and human performance. *Acta Psychologica*, *53*(1), 61–97. https://doi.org/10.1016/0001-6918(83)90016-1

Sanders, A. F. (2013). *Elements of human performance: Reaction processes and attention in human skill.* Hoboken: Taylor and Francis.

Sarter, N. B. (2008). Investigating mode errors on automated flight decks: Illustrating the problem-driven, cumulative, and interdisciplinary nature of human factors research. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *50*(3), 506–510. https://doi.org/10.1518/001872008X312233

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *37*(1), 5–19. https://doi.org/10.1518/001872095779049516

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1926–1943). New York: Wiley.

Sauer, J., Chavaillaz, A., & Wastell, D. (2015). Experience of automation failures in training: Effects on trust, automation bias, complacency, and performance. *Ergonomics*, 1–28. https://doi.org/10.1080/00140139.2015.1094577

Saxby, D. J., Matthews, G., Hitchcock, E. M., Warm, J. S., Funke, G. J., & Gantzer, T. (2008). Effect of active and passive fatigue on performance using a driving simulator. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(21), 1751–1755. https://doi.org/10.1177/154193120805202113

Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., & Neubauer, C. (2013). Active and passive fatigue in simulated driving: Discriminating styles of workload regulation and their safety impacts. *Journal of Experimental Psychology: Applied*, *19*(4), 287–300. https://doi.org/10.1037/a0034386

Scarpello, V., & Campbell, J. P. (1983). Job satisfaction: Are all the parts there? *Personnel Psychology*, *36*(3), 577–600. https://doi.org/10.1111/j.1744-6570.1983.tb02236.x

Scerbo, M. W. (1998). What's so boring about vigilance? In R. R. Hoffman, M. F. Sherrick, & J. S. Warm (Eds.), *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 145–166). Washington: American Psychological Association. https://doi.org/10.1037/10290-006

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *58*(3), 377–400. https://doi.org/10.1177/0018720816634228

Schlag, B. (2013). Persönliche Veränderungen der Mobilität und der Leistungsfähigkeit im Alter [Individual changes in mobility and performance in old age]. In B. Schlag & K. J. Beckmann (Eds.), *Mobilität und Alter: Vol. 7. Mobilität und demografische Entwicklung* (pp. 119–144). Köln: TÜV Media.

Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., & Kincses, W. E. (2009). Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accident Analysis & Prevention*, *41*(5), 1087–1093. https://doi.org/10.1016/j.aap.2009.06.007

Schmidt, J., Braunagel, C., Stolzmann, W., & Karrer-Gauss, K. (2016). Driver drowsiness and behavior detection in prolonged conditionally automated drives. In *2016 IEEE Intelligent Vehicles Symposium (IV)* (pp. 400–405). https://doi.org/10.1109/IVS.2016.7535417

Schneider, W., & Detweiler, M. (1988). The role of practice in dual-task performance: Toward workload modeling in a connectionist/control architecture. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *30*(5), 539–566. https://doi.org/10.1177/001872088803000502

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. https://doi.org/10.1037/0033-295X.84.1.1

Schoettle, B., & Sivak, M. (2014a). *A survey of public opinion about autonomous and self-driving vehicles in the U.S., The U.K., and Australia* (UMTRI No. 2014-21). Ann Arbor, MI.

Schoettle, B., & Sivak, M. (2014b). *Public opinion about self-driving vehicles in China, India, Japan, the U.S., the U.K., and Australia* (No. UMTRI-2014-30). Ann Arbor, MI.

Schoettle, B., & Sivak, M. (2015). *Motorists' preferences for different levels of vehicle automation: 2015* (UMTRI No. 2015-22). Ann Arbor, MI.

Schoettle, B., & Sivak, M. (2016). *Motorists' preferences for different levels of vehicle automation: 2016* (No. SWT-2016-8). Ann Arbor, MI.

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339. https://doi.org/10.1037/met0000061

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of Management Review*, *32*(2), 344–354. https://doi.org/10.5465/AMR.2007.24348410

Schwebel, D. C., Ball, K. K., Severson, J., Barton, B. K., Rizzo, M., & Viamonte, S. M. (2007). Individual difference factors in risky driving among older adults. *Journal of Safety Research*, *38*(5), 501–509. https://doi.org/10.1016/j.jsr.2007.04.005

Seli, P., Cheyne, J. A., & Smilek, D. (2013). Wandering minds and wavering rhythms: Linking mind wandering and behavioral variability. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 1–5. https://doi.org/10.1037/a0030954

Seppelt, B. D., & Lee, J. D. (2007). Making adaptive cruise control (ACC) limits visible. *International Journal of Human-Computer Studies*, *65*(3), 192–205. https://doi.org/10.1016/j.ijhcs.2006.10.001

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth Cengage Learning.

Shaw, T. H., Matthews, G., Warm, J. S., Finomore, V. S., Silverman, L., & Costa, P. T. (2010). Individual differences in vigilance: Personality, ability and states of stress. *Journal of Research in Personality*, *44*(3), 297–308. https://doi.org/10.1016/j.jrp.2010.02.007

Shergold, I., Wilson, M., & Parkhurst, G. (2016). *The mobility of older people, and the future role of connected autonomous vehicles* (Project Report). Bristol. Retrieved from Centre for Transport and Society, University of the West of England website: http://eprints.uwe.ac.uk/31998

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. Cambridge, MA: MIT Press.

Sheridan, T. B. (2002). *Humans and automation: System design and research issues*. *Wiley series in system engineering and management: Vol. 3*. Hoboken, N.J.: Wiley.

Sheridan, T. B., & Parasuraman, R. (2005). Human-automation interaction. *Reviews of Human Factors and Ergonomics*, *1*(1), 89–129. https://doi.org/10.1518/155723405783703082

Shinar, D., Tractinsky, N., & Compton, R. (2005). Effects of practice, age, and task demands, on interference from a phone task while driving. *Accident Analysis & Prevention*, *37*(2), 315–326. https://doi.org/10.1016/j.aap.2004.09.007

Shinoda, H., Hayhoe, M. M., & Shrivastava, A. (2001). What controls attention in natural environments? *Vision Research*, *41*(25-26), 3535–3545. https://doi.org/10.1016/S0042-6989(01)00199-7

Sidman, M. (1952). A note on functional relations obtained from group data. *Psychological Bulletin*, *49*(3), 263–269. https://doi.org/10.1037/h0063643

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*(3), 300–304. https://doi.org/10.1037/h0020586

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074. https://doi.org/10.1068/p2952

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*(1), 16–20. https://doi.org/10.1016/j.tics.2004.11.006

Singh, I. L., Molloy, R., Mouloua, M., Deaton, J., & Parasuraman, R. (1998). Cognitive ergonomics of cockpit automation. In I. L. Singh & R. Parasuraman (Eds.), *Human cognition: A multidisciplinary perspective* (pp. 242–253). New Delhi, Thousand Oaks, CA: Sage Publications.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993a). Automation-induced "complacency": Development of the complacency-potential rating scale. *The International Journal of Aviation Psychology*, *3*(2), 111–122. https://doi.org/10.1207/s15327108ijap0302_2

Singh, I. L., Molloy, R., & Parasuraman, R. (1993b). Individual differences in monitoring failures of automation. *The Journal of General Psychology*, *120*(3), 357–373. https://doi.org/10.1080/00221309.1993.9711153

Singh, S. (2015). *Critical reasons for crashes investigated in the national motor vehicle crash causation survey* (DOT HS No. 812 115). Washington, D. C.

Sitkin, S. B., & Pablo, A. L. (1992). Reconceptualizing the determinants of risk behavior. *Academy of Management Review*, *17*(1), 9–38. https://doi.org/10.5465/AMR.1992.4279564

Siu, K.-C., Chou, L.-S., Mayr, U., van Donkelaar, P., & Woollacott, M. H. (2008). Does inability to allocate attention contribute to balance constraints during gait in older adults? *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *63*(12), 1364–1369.

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, *51*(5), 991–1006. https://doi.org/10.1006/ijhc.1999.0252

Sloan, J. A., Aaronson, N., Cappelleri, J. C., Fairclough, D. L., & Varricchio, C. (2002). Assessing the clinical significance of single items relative to summated scores. *Mayo Clinic Proceedings*, *77*(5), 479–487. https://doi.org/10.4065/77.5.479

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*(6), 946–958. https://doi.org/10.1037/0033-2909.132.6.946

Sohn, S. Y., & Stepleman, R. (1998). Meta-analysis on total braking time. *Ergonomics*, *41*(8), 1129–1140. https://doi.org/10.1080/001401398186432

Spain, R. D., Bustamante, E. A., & Bliss, J. P. (2008). Towards an empirically developed scale for system trust: Take two. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*(19), 1335–1339. https://doi.org/10.1177/154193120805201907

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149.

Stanton, N. A., & Young, M. S. (1998). Vehicle automation and driving performance. *Ergonomics*, *41*(7), 1014–1028. https://doi.org/10.1080/001401398186568

Sternberg, R. J., Sternberg, K., & Mio, J. (2012). *Cognitive psychology* (6th ed.). Belmont, Calif.: Wadsworth.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315.

Stevenson, R. (2011). Long-distance car radar. *IEEE Spectrum*.

Straussberger, S. (2006). Monotony in air traffic control – Contributing factors and mitigation strategies (Doctoral dissertation). Karl-Franzens-University Graz, Graz.

Stutts, J. C., Reinfurt, D. W., Staplin, L., & Rodgman, E. A. (2001). *The role of driver distraction in traffic crashes*. Washington, D. C.

Szucs, D., & Ioannidis, J. P. A. (2017). When null hypothesis significance testing is unsuitable for research: A reassessment. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00390

Tal, E. (2017). Measurement in science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University.

Temple, J. G., Warm, J. S., Dember, W. N., Jones, K. S., LaGrange, C. M., & Matthews, G. (2000). The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *42*(2), 183–194. https://doi.org/10.1518/001872000779656480

Thiffault, P., & Bergeron, J. (2003). Monotony of road environment and driver fatigue: A simulator study. *Accident Analysis & Prevention*, *35*(3), 381–391.

Thompson, C. (2015, November 4). Elon Musk: In less than 20 years, owning a car will be like owning a horse. *Business Insider*. Retrieved from http://www.businessinsider.com/elon-musk-owning-a-car-in-20-years-like-owning-a-horse-2015-11

Tijerina, L., Parmer, E., & Goodman, M. J. (1998). Driver workload assessment of route guidance system destination entry while driving: A test track study. In *Proceedings of the 5th ITS World Congress* (pp. 12–16).

Törnros, J., & Bolling, A. (2006). Mobile phone use – effects of conversation on mental workload and driving speed in rural and urban environments. *Transportation Research Part F: Traffic Psychology and Behaviour*, *9*(4), 298–306. https://doi.org/10.1016/j.trf.2006.01.008

Treisman, A. M. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, *31*(2), 156–177. https://doi.org/10.1016/S0734-189X(85)80004-9

Twisk, D. A. M., & Stacey, C. (2007). Trends in young driver risk and countermeasures in European countries. *Journal of Safety Research*, *38*(2), 245–257. https://doi.org/10.1016/j.jsr.2007.03.006

Uebersax, J. S. (2006). Likert scales: Dispelling the confusion. Retrieved from http://www.john-uebersax.com/stat/likert.htm

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., & Crundall, D. (2003). Visual attention while driving: Sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, *46*(6), 629–646. https://doi.org/10.1080/0014013031000090116

Underwood, G., Crundall, D., & Chapman, P. (2002). Selective searching while driving: The role of experience in hazard detection and general surveillance. *Ergonomics*, *45*(1), 1–12. https://doi.org/10.1080/00140130110110610

VandenBos, G. R. (Ed.). (2015). *APA dictionary of psychology* (2nd ed.). Washington, DC: American Psychological Association.

Verberne, F. M. F., Ham, J., & Midden, C. J. H. (2012). Trust in smart systems: Sharing driving goals and giving information to increase trustworthiness and acceptability of smart systems in cars. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *54*(5), 799–810. https://doi.org/10.1177/0018720812443825

Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin*, *122*(3), 231–249. https://doi.org/10.1037/0033-2909.122.3.231

Visser, E., Pijl, Y. J., Stolk, R. P., Neeleman, J., & Rosmalen, J. G. M. (2007). Accident proneness, does it exist? A review and meta-analysis. *Accident Analysis & Prevention*, *39*(3), 556–564. https://doi.org/10.1016/j.aap.2006.09.012

De Visser, E., Shaw, T., Mohamed-Ameen, A., & Parasuraman, R. (2010). Modeling human-automation team performance in networked systems: Individual differences in working memory count. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(14), 1087–1091. https://doi.org/10.1177/154193121005401408

De Waard, D. (1996). The measurement of drivers' mental workload (Doctoral dissertation). University of Groningen, Haren, Netherlands.

De Waard, D., van der Hulst, M., Hoedemaeker, M., & Brookhuis, K. (1999). Driver behavior in an emergency situation in the automated highway system. *Transportation Human Factors*, *1*(1), 67–82. https://doi.org/10.1207/sthf0101_7

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem

(2011). *Journal of Personality and Social Psychology*, *100*(3), 426–432. https://doi.org/10.1037/a0022790

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Wagner, M., & Koopman, P. (2015). A philosophy for developing trust in self-driving cars. In G. Meyer & S. Beiker (Eds.), *Lecture notes in mobility. Road vehicle automation 2* (pp. 163–171). Cham: Springer International Publishing.

Walker, E., & Nowacki, A. S. (2011). Understanding equivalence and noninferiority testing. *Journal of General Internal Medicine*, *26*(2), 192–196. https://doi.org/10.1007/s11606-010-1513-8

Walker, G. H., Stanton, N. A., & Salmon, P. (2016). Trust in vehicle technology. *International Journal of Vehicle Design*, *70*(2), 157. https://doi.org/10.1504/IJVD.2016.074419

Warshawsky-Livne, L., & Shinar, D. (2002). Effects of uncertainty, transmission type, driver age and gender on brake reaction and movement time. *Journal of Safety Research*, *33*(1), 117–128.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Watson, J. M., & Strayer, D. L. (2010). Supertaskers: Profiles in extraordinary multitasking ability. *Psychonomic Bulletin & Review*, *17*(4), 479–485. https://doi.org/10.3758/PBR.17.4.479

Watzenig, D., & Horn, M. (Eds.). (2017a). *Automated driving: Safer and more efficient future driving*. Cham, s.l.: Springer International Publishing.

Watzenig, D., & Horn, M. (2017b). Introduction to automated driving. In D. Watzenig & M. Horn (Eds.), *Automated driving: Safer and more efficient future driving* (pp. 3–16). Cham, s.l.: Springer International Publishing.

Weller, G., Heyne, F., Feige, T., Bretschneider, H., Oeser, H., & Schlag, B. (2011). Die Wirkung gerichteter Warnungen von Fahrerassistenzsystemen auf die Blickzuwendungs- und Reaktionszeiten von Autofahrern [The impact of directed warnings of driver assistance systems on drivers' gaze behavior and reaction times]. In S. Schmid, M. Elepfandt, J. Adenauer, & A., Lichtenstein (Eds.): *Vol. 33. Berichte aus dem Zentrum Mensch-Maschine-Systeme der Technischen Universität Berlin, Reflexionen und Visionen der Mensch-Maschine-Interaktion - aus der Vergangenheit lernen, Zukunft gestalten: 9. Berliner Werkstatt Mensch-Maschine-Systeme* (pp. 372–378). Düsseldorf: VDI-Verlag.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. https://doi.org/10.1080/14639220210123806

Wickens, C. D., & Carswell, C. M. (2006). Information processing. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 111–149). Hoboken, NJ, USA: John Wiley & Sons. https://doi.org/10.1002/0470048204.ch5

Wickens, C. D., Clegg, B. A., Vieane, A. Z., & Sebok, A. L. (2015). Complacency and automation bias in the use of imperfect automation. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *57*(5), 728–739. https://doi.org/10.1177/0018720815581940

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2016). *Engineering psychology and human performance* (4th ed.). London, New York: Routledge Taylor & Francis Group.

Wickens, C. D., Lee, J. D., Liu, Y., & Gordon-Becker, S. (2014). *An introduction to human factors engineering* (2nd ed.). *Always learning*. Harlow: Pearson Education.

Wickens, C. D., Li, H., Santamaria, A., Sebok, A., & Sarter, N. B. (2010). Stages and levels of automation: An integrated meta-analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(4), 389–393. https://doi.org/10.1177/154193121005400425

Wickens, C. D., & McCarley, J. S. (2008). *Applied attention theory*. Boca Raton, Fla.: CRC Press.

Wickens, C. D., & Xu, X. (2002). *Automation trust, reliability and attention* (Technical Report No. AHFD-02-14/MAAD-02-2). Savoy, Illinois.

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, *23*(10), 995–1011. https://doi.org/10.1080/00140138008924809

Wiggins, M., Stevens, C., Howard, A., Henley, I., & O'Hare, D. (2002). Expert, intermediate and novice performance during simulated pre-flight decision-making. *Australian Journal of Psychology*, *54*(3), 162–167. https://doi.org/10.1080/00049530412331312744

Wilde, G. J. S. (1982). The theory of risk homeostasis: Implications for safety and health. *Risk Analysis*, *2*(4), 209–225. https://doi.org/10.1111/j.1539-6924.1982.tb01384.x

Winner, H., Hakuli, S., Lotz, F., & Singer, C. (Eds.). (2016). *Handbook of driver assistance systems: Basic information, components and systems for active safety and comfort.* Cham: Springer Reference.

De Winter, J. C. F., Leeuwen, P. M. van, & Happee, R. (2012). Advantages and disadvantages of driving simulators: A discussion. In A. Spink, F. Grieco, O. Krips, L. Loijens, L. Noldus, & P. Zimmerman (Eds.), *Proceedings of Measuring Behavior 2012: 8th International Conference on Methods and Techniques in Behavioral Research* (47–50). Wageningen: Noldus Information Technology bv.

Wood, G., Hartley, G., Furley, P. A., & Wilson, M. R. (2016). Working memory capacity, visual attention and hazard perception in driving. *Journal of Applied Research in Memory and Cognition*, *5*(4), 454–462. https://doi.org/10.1016/j.jarmac.2016.04.009

Yang, X., Rakheja, S., & Stiharu, I. (2001). Adapting an articulated vehicle to its drivers. *Journal of Mechanical Design*, *123*(1), 132. https://doi.org/10.1115/1.1336797

Yanko, M. R., & Spalek, T. M. (2013). Driving with the wandering mind: The effect that mind-wandering has on driving performance. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *56*(2), 260–269. https://doi.org/10.1177/0018720813495280

Young, M. S., & Stanton, N. A. (2002). Malleable Attentional Resources Theory: A new explanation for the effects of mental underload on performance. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *44*(3), 365–375. https://doi.org/10.1518/0018720024497709

Zedeck, S. (Ed.). (2014). *APA dictionary of statistics and research methods* (1st ed.). Washington, DC: American Psychological Association.

Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accident Analysis & Prevention*, *78*, 212–221. https://doi.org/10.1016/j.aap.2015.02.023

Appendix

A Complementary articles and methods

Besides the articles that are included in this thesis, further research has been conducted. The following is a selection of work that shines a light on facets that are not in the main focus of this thesis or that has laid the methodical foundations of the included articles.

A1 "Why do I have to drive now? Post hoc explanations of take-over requests"

Körber, M., Prasch, L., & Bengler, K. (2018). Why do I have to drive now? Post hoc explanations of takeover requests. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *60*(3), 305–323. https://doi.org/10.1177/0018720817747730

In this work, we conducted further research on trust in automation and investigated if trust in automation and acceptance can be increased by providing an explanation for take-over requests. The concept of automated driving is a novelty for the majority of the driving population, which is why its acceptance is not directly guaranteed but has to be ensured (Payre et al., 2014). Whereas flawless system performance may be sufficient from a technical point of view, a driving automation system's functionality must be understood and valued by the drivers to establish its acceptance and use (Adell et al., 2014; van der Laan, Heino, & de Waard, 1997). According to Ghazizadeh et al. (2012), trust determines how the operator's beliefs and external variables affect the perceived usefulness and ease of use of an automated vehicle. It also has a direct effect on the behavioral intention to use automation. Thus, trust in automation is a necessary condition that has to be established before a system will be accepted. In comparison to ADAS, a driving automation system is a more sophisticated technology, which entails an increase in autonomy and authority (Parasuraman et al., 2000). Such a complex machine agent requires more coordination than a simple ADAS function and an adequate model of the automated system's intentions and actions (Norman, 1990; Sarter, 2008). Otherwise, the increase in autonomy and authority creates "mysterious and obstinate black boxes" (Christoffersen & Woods, 2002, p. 4). Thus, to ensure trust in automation, it is crucial to provide drivers with comprehensible information regarding the automated system's intentions, state, and capacity in order to help them understand and to make the system predictable. Promoting predictability, understanding, or transparency has been shown to increase trust in automation in several studies (Forster, Naujoks, & Neukum, 2017; Hoff & Bashir, 2015; Verberne, Ham, & Midden, 2012). An explanation also helps to prevent automation surprises (Sarter et al., 1997) and concomitant negative emotional reactions, which are known to reduce acceptance. In this context, we investigated whether providing an explanation for a take-over request in automated driving increases understanding and system transparency and, in doing so, trust in automation and acceptance.

Forty participants were equally assigned to either an experimental group, which received an explanation of the reason for a take-over request, or a control group without explanations. In a simulator drive, both groups experienced three take-over situations that varied in the obviousness of the cause for the take-over request. The participants rated their trust in automation before and after each take-over situation and rated their acceptance before and after the experimental drive. All participants reported the same high level of acceptance before and after the drive and we found no evidence for a difference between both groups. Trust ratings remained unchanged by take-overs in all but a single situation. Participants provided with explanation felt stronger that they had understood the system and the reasons for the take-overs. While the explanations had no systematic effect on trust in automation or acceptance, the increase in system transparency by providing explanations seems to have been successful.

A2 "Vigilance decrement and passive fatigue caused by monotony in automated driving"

Körber, M., Cingel, A., Zimmermann, M., & Bengler, K. (2015). Vigilance decrement and passive fatigue caused by monotony in automated driving. *Procedia Manufacturing*, *3*, 2403–2409. https://doi.org/10.1016/j.promfg.2015.07.499

The model of May and Baldwin (2009) describes passive fatigue as the opposite of active fatigue. Unlike active engagement, task underload and monotony are seen as the cause of fatigue effects such as a reduced attentional capacity or increased reaction times (Matthews & Desmond, 2002; Saxby, Matthews, Warm, Hitchcock, & Neubauer, 2013; Young & Stanton, 2002). Schmidt et al. (2009) empirically investigated the effect of passive fatigue in a naturalistic setting. In their study, they examined the effect of a 3 h monotonous highway drive on driver's vigilance and drowsiness. Besides an increase in subjective and objective indicators of drowsiness, reaction times increased during the drive. The passive monitoring role in partially automated driving (Level 2; SAE International, 2016) could constitute an even more monotonous situation since the driver's only task is to monitor the system and to intervene in case of an event. In this work, we investigated the decrement of vigilance during a partially automated highway drive with 20 participants driving in a driving simulator for 42 min. Indicators for the vigilance state were a self-developed auditory oddball paradigm, eye tracking, and a mind wandering questionnaire (Matthews, Joyner, Gilliland, Huggins, & Falconer, 1999). We found no significant effects of time-on-task on the reaction times in the oddball task but a significant increase of drowsiness indicated by three of four eye tracking measures (blink frequency, blink duration, pupil diameter). Also, subjectively reported mind wandering increased in the course of the drive. Altogether, the results show that fatigue can occur even without active task engagement. Yet, future studies have to clarify its consequences on reaction time and reaction quality in critical events.

A3 Articles on statistical methods

Körber, M., Radlmayr, J., & Bengler, K. (2016). Bayesian highest density intervals of take-over times for highly automated driving in different traffic densities. In *Proceedings of the Human Factors and Ergonomics Society (HFES) 60th Annual Meeting 2016* (Vol. 60, pp. 2009–2013). https://doi.org/10.1177/1541931213601457

Radlmayr, J., Körber, M., Feldhütter, A., & Bengler, K. (2016). Methoden und Fahrermodelle für Hochautomatisiertes Fahren [Methods and driver models for highly automated driving]. In K. Kompaß (Ed.), *Haus der Technik Fachbuch: Band 144. Methodenentwicklung für aktive Sicherheit und automatisiertes Fahren. 2. Expertendialog zu Wirksamkeit - Beherrschbarkeit - Absicherung.* Renningen: expert Verlag.

Körber, M. (2016). Einführung in die inferenzstatistische Auswertung mit Bayes-Statistik [Introduction to inferential statistics with Bayesian statistics]. *Ergonomie aktuell*, 017, 27–33.

Besides empirical studies in the field of human factors, several articles on statistical methods have been published in the context of this dissertation. These articles describe the rationale for the advancement in statistical methods used in the articles of this thesis. While the early articles, Article 3 and 5, solely rely on NHST, Article 2 constitutes a shift to more informative ways of analysis in form of effect size estimation and interval estimation. Reliance on NHST promotes a dichotomous in data analysis: significant result or non-significant result (Szucs & Ioannidis, 2017). This practice is especially uninformative if a nil hypothesis of zero difference is tested. Moving on to estimating the magnitude of an effect (raw or standardized) promotes a focus on the actual effects obtained and the precision of the estimates (Cumming, 2014). Article 4 represents a switch to Bayesian data analysis. Here, we relied on Bayesian parameter estimation to quantify the uncertainty in the parameter estimates. For hypothesis testing, we relied on BFs, which are defined as the ratio of the probability of the data given a null model to the probability of the data given an alternative model (Kruschke, 2015; Rouder et al., 2009). The ratio, thus, quantifies whether the data are more compatible with the null or the alternative model and provide a continuous form of evidence either for the null hypothesis or the alternative hypothesis (or neither; Dienes, 2014; Schönbrodt et al., 2017).

A4 Theoretical considerations and the development of a questionnaire to measure trust in automation[2]

An empirical investigation of the role of trust in automation in automated driving necessitates a measurement of trust in automation. Trust in automation is a latent construct, which is not directly observable; thereby, researchers rely on indicators such as neuroscientific methods (Drnec et al., 2016), behavioral measures (e.g., eye tracking; Hergeth et al., 2016), or questionnaires (Jian, Bisantz, & Drury, 2000; Madsen & Gregor, 2000). Trust in automation and reliance on automation are closely related: "People tend to rely on automation they trust and tend to reject automation they do not" (Lee & See, 2004, p. 51). Yet, trust in automation and reliance on automation are at the same time distinct constructs. In their theory of reasoned action, Ajzen and Fishbein (1980) argue that behavior, such as reliance, results from an intention and that this intention is a function of attitudes, which in turn are an affective evaluation of beliefs. Trust in automation as an attitude, thus, stands between the belief about the characteristics of an automated system, such as its reliability, and the intention to rely on it. Attitude, intention, and actual behavior are not in a deterministic but in a probabilistic relationship (Ajzen & Fishbein, 1980). Whether trust translates into reliance behavior depends on a dynamic interaction of operator, automation, situational factors, and interface (Lee & See, 2004). As a result, other factors, such as the effort to engage or self-confidence, also affect the intention to rely on an automated system (Bisantz & Seong, 2001; Dzindolet, Beck, Pierce, & Dawe, 2001; Kirlik, 1993; Lee & See, 2004; Meyer, 2004). Environmental and cognitive constraints, such as time pressure, then determine whether a formed intention translates into actual reliance on automation. Even if trust is at a high level and the automated system is perceived as capable, reliance does not necessarily follow (Kirlik, 1993). Figure 16 from Lee and See (2004) provides a notable overview of the relationship between trust in automation and reliance. That means, to measure trust as an attitude itself, a questionnaire or another similar methodology that is distinct from observable risk taking is necessary (Mayer et al., 1995). Furthermore, the conceptualizations of trust in automation refer to the construct as an attitude (Lee & See, 2004), a mainly affective response closely related to beliefs and expectations. Affective responses are not always accompanied by overt behavior. For example, students with and without math anxiety may behave the same way during a math test even though their internal state differs (McCoach, Gable, & Madura, 2013). An affective response is, thereby, probably only completely accessible through self-report (Paulhus & Vazire, 2007). A questionnaire, therefore, is an attractive method to measure trust in automation.

[2] Parts of this section have been published in Körber (2019).

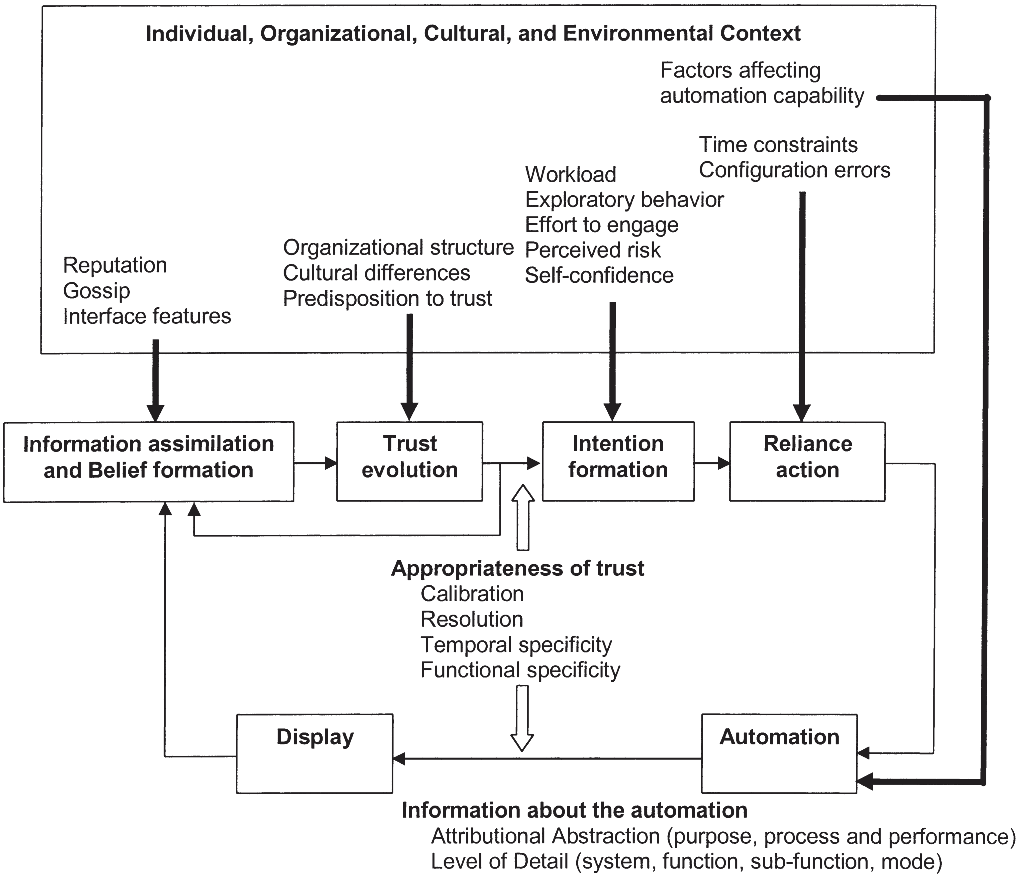

Figure 16. The model of the evolution of trust and its relationship to reliance from Lee and See (2004).

A literature review of available questionnaires on trust in automation revealed that the questionnaires comprise single-item as well as multi-item scales. Single-item scales allow a quick, uncomplicated measurement such as a dynamic assessment during an experiment. However, these instruments also have some drawbacks. Dimensions and models of trust have been extensively discussed, resulting in a variety of facets and concepts (Lee & See, 2004). It is questionable whether the broadness and depth of this construct can be captured by a single questionnaire item. In contrast, multiple, heterogeneous indicators (= questionnaire items) enhance construct validity by increasing the probability of adequately identifying the construct (Eisinga, Grotenhuis, & Pelzer, 2013). The items of a scale should hold a common core but also contribute unique variance not shared and untapped by other items (Churchill & Peter, 1984). Consequently, Fuchs and Diamantopoulos (2009) do not recommend single-item scales if the construct in question is abstract. Likewise, a single item does not allow for a detailed analysis of the underlying reasons for a favorable or non-favorable trust score. Is the machine perceived as unreliable? Does a participant simply not trust a certain brand? It is not possible to give an answer with a single item scale.

Furthermore, single-item scales are more vulnerable to random measurement error because random errors cannot be canceled out by the remaining items like in a multi-item scale (Emons, Sijtsma, & Meijer, 2007). The first axiom of classical test theory (CTT) claims that the true score τ_{vi} can be estimated by the expected value of multiple measurements of a participant v of item i: $\tau_{vi} = E(x_{vi})$. Given (essentially) tau-equivalent items, each of these on their own is a measurement of the latent construct. The second axiom states that the obtained score, i.e. the item responses, which are meant to represent the latent construct, of a participant v at item i, x_{vi}, is a combination of the true score τ_{vi} and a random measurement error ε_{vi} (Zedeck, 2014). Thus, with multiple measurements, random measurement error occurring at each measurement is canceled out and the true score is the expected value of the multiple measurements. In this way, the measurement gains precision with each added item. In contrast, the estimation of the true score is not possible with just a single item. Shorter tests possess a lower reliability because of higher measurement error (Bühner, 2011; DeVellis, 2006; Moosbrugger & Kelava, 2012; Robins, Hendin, & Trzesniewski, 2001).

Moreover, the items of a questionnaire generally represent a random selection from all hypothetically possible indicators of the construct (Nunnally & Bernstein, 1994). Using multiple items helps to cancel out errors due to specificities inherent in single items (Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012). Single-item scales are correspondingly more susceptible to unknown biases in meaning and interpretation (Hoeppner, Kelly, Urbanoski, & Slaymaker, 2011): If a single item is misunderstood, the validity of the whole measurement becomes questionable. Due to respondents' state dependence, an individual's response pattern frequently carries over to the subsequent item. This is less problematic in multi-item scales because the remaining items can possibly compensate for this effect. If a scale consists of only one item, such carry-over effects can skew the response, lowering a measure's (predictive) validity (Diamantopoulos et al., 2012). Furthermore, the estimation of the reliability, i.e. the precision of a measurement, is problematic with only a single item. Reliability is defined as the ratio of the variance of the true scores to the variance of the observed scores. Without measurement error, all variation in observed scores would be explained by variation in true scores. With a single item, it is not possible to calculate the variance of the true scores and, consequently, the measurement error and confidence intervals for a measurement as well. If items are (essentially) tau-equivalent, then the correlation between them can be used to estimate the reliability by calculating the internal consistency, Cronbach's α (Bühner, 2011). With single items, however, the internal consistency reliability statistic cannot be computed (Hoeppner et al., 2011). That means the precision of a measurement remains unknown. Single-item scales have to use other forms of reliability estimation that might have drawbacks in the context of trust measurement. If the typical data analysis approach based on CTT is followed, multi-item scales offer the advantage that averaging over multiple items allows for a more granular range of possible resulting scores, which can be seen as an argument for measurement (of a continuous construct) on an interval scale level. Also, typically less skewed data is produced (Carifio & Perla, 2007). Because of these drawbacks, single item measures were discarded as an option to measure trust in the experiment.

Jian et al. (2000) constructed a 12-item questionnaire following an inductive approach (Bühner, 2011) not based on the hitherto existing theoretical models of trust. Seven graduate students majoring in Linguistics or English with unknown affinity to technology rated 138 words regarding their relation to general trust, human-human trust, and human-machine trust. No definition of these three constructs was given, but each student used his own definition, which were also used for word collection. It is unclear on which definition of trust in automation the questionnaire is based or what construct it exactly claims to measure. Not reporting relevant statistics, the authors conducted a factor analysis and a cluster analysis on the relatedness of the generated words without evaluating the questionnaire in the interaction with an existing automated system. Items were then generated based on the emerged groups. Because ratings of trust were highly negatively correlated with ratings of distrust, the authors proclaim a unidimensional structure of trust. This structure has been questioned in a more recent analysis by Spain, Bustamante, and Bliss (2008), who used a not further specified "modified version" (p. 1337), and in the thorough analysis of their German adaptation by Pöhler, Heine, and Deml (2016). Following this notion, an individual could trust and mistrust an automated system at the same time. Nonetheless, others have found support for the assumption that trust and distrust form one dimension (Onnasch, Wiczorek, & Manzey, 2011). For example, the probably most cited definition of trust defines the construct as "the willingness of a party to be vulnerable to the actions of another party" (Mayer et al., 1995, p. 712). Following this definition, an individual would not be willing to be vulnerable and take risks at all if trust is completely absent (Schoorman, Mayer, & Davis, 2007). Both a complete absence of trust and distrust would mean the same thing. Regardless of dimensionality, it is questionable how well terms like *friendship*, *honor, love*, or *steal* can be evaluated in relation to an automated system. This content is probably included because of the process of item generation.

The authors generated the questionnaire items based on terms that were collected at the same time regarding their assumed affinity to general trust, human-human trust, and human-automation trust. This led to the inclusion of terms that are probably more relevant to interpersonal trust and that require a malicious intent (e.g., *phony*, *honor*). Since the authors generated the item content based on these terms, the items also focus on this specific facet, for example deception. However, despite some similarities, interpersonal trust and trust in automation exhibit fundamental differences in their basis, development, and in the relationship between trustor and trustee (Hoff & Bashir, 2015; Madhavan & Wiegmann, 2007). One fundamental difference is that automation lacks intentionality (Lee & See, 2004). Although designers of automated systems develop it with a certain purpose, deception requires a malicious intent, which does not seem to fit perfectly in the context of thoroughly developed automated systems. The items mapped to terms related to distrust (item 1 to 5) seem suitable for interpersonal trust measurement but may not completely match the context of human-automation trust. As a result, this lowers construct validity because of fuzzy construct boundaries. Moreover, albeit trust is assumed to be unidimensional by Jian et al. (2000), almost half of the items in their questionnaire are inversely/negatively formulated and all items are sorted by their relation to trust (items 1 to 5 negative, items 6 to 12 positive). This promotes item order effects (Elson, 2017), which might be another explanation for the better fit of a two-factor solution found by Pöhler et al. (2016). Given its focus on malicious intent, it was decided to discard this questionnaire. Another multi-item (25 items) questionnaire by Madsen and Gregor (2000) focuses on human-computer trust. However, their initial study did not support their theoretical model, which also includes the general appraisal of the system. For these reasons, this questionnaire was also discarded. As a consequence, it was decided to develop an entirely new questionnaire.

The measurement of a latent construct such as trust requires the process of construct validation (Flake, Pek, & Hehman, 2017). In the substantive phase, the literature is reviewed, the construct is defined and conceptualized, and its dimensions, boundaries, and structure are identified. For this purpose, theoretical discourses on trust in automation were screened along with empirical articles and articles with a stronger focus on interpersonal trust[3]. The most widespread and most cited model of trust is the dyadic model of organizational trust by Mayer et al. (1995), depicted in Figure 17. Integrating previous theoretical accounts on trust, the parsimonious model differentiates trust from its contributing factors and its outcome, risk taking in a relationship. The authors argue that trust is only necessary in a risky situation or when having something invested. In this context, they define trust as

> the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party. (Mayer et al., 1995, p. 712)

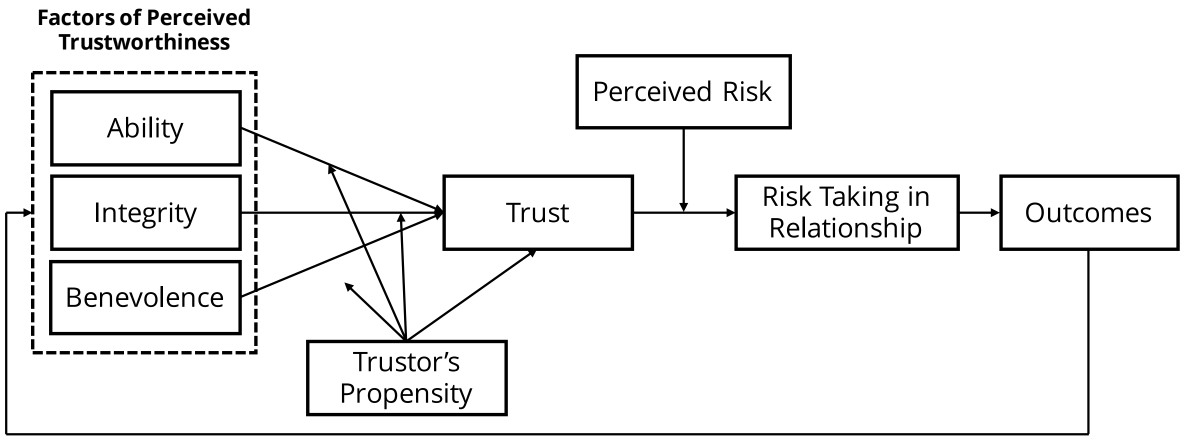

Figure 17. Illustration of the trust model from Mayer et al. (1995).

According to their model, a person's trust depends on two components, a person's *individual propensity* or general willingness to trust others and the *trustworthiness* of the party to be trusted (trustee). A person's trust propensity results from different developmental experiences, personality type, and cultural background and determines how much a person trusts a trustee prior to any knowledge of that particular party being available. The second component, the perceived trustworthiness, is determined by three relevant attributes of the trustee: 1) *Ability*: The level of skills, competencies, and characteristics that the trustee possesses and that enables him to have influence within a specific domain. 2) *Benevolence*: The extent to which a trustee is perceived to want to do good to the trustor and avoids egocentric motives. 3) *Integrity*: The extent to which the trustee

[3] In this literature review, the following work was considered: Barber (1983), Blomqvist (1997), Butler and Cantrell (1984), Butler (1991), Deutsch (1958), Deutsch (1960), Dzindolet et al. (2001), Hoff and Bashir (2015), Hoffman, Johnson, Bradshaw, and Underbrink (2013), Jian et al. (2000), Lee and Moray (1992), Lee and See (2004), Madhavan and Wiegmann (2007), Madsen and Gregor (2000), Mayer et al. (1995), McKnight and Chervany (1996), McKnight and Chervany (2001), Muir (1987), Muir (1994), Muir and Moray (1996), Rempel, Holmes, and Zanna (1985), Rotter (1971).

consistently adheres to a series of principles that the trustor deems acceptable. Risk taking is then the behavioral manifestation of the willingness to be vulnerable, i.e. the outcome of trust.

The increasing interaction with automated systems has sparked the interest of human factors researchers in trust in automation with the overall goal to improve joint system performance in mind (Drnec et al., 2016). Since interpersonal trust and trust in automation, as already mentioned, exhibit fundamental differences, the model from Mayer et al. (1995) does not completely apply to trust in automation. Taking this into account, Lee and See (2004) follow the model of trust by Mayer et al. (1995) but fit their dimensions to the context of trust in automation. They argue that previously found bases for trust in automation can be summarized into three dimensions, *performance*, *process*, and *purpose*, which correspond to the dimensions of trustworthiness in the model by Mayer et al. (1995), as illustrated in Figure 18. *Performance* refers to the current and previous operation of the automated system and comprises characteristics such as reliability, competency, and ability. Performance information describes what the automated system can do reliably and matches the attribute ability in Mayer et al. (1995). *Process* describes how the automated system operates and if this modus operandi is appropriate for the situation and the operator's goals. It subsumes characteristics such as understandability and matches integrity in Mayer et al. (1995). *Purpose* describes the intention in the automated system's design, the perception that the designers possess a positive orientation towards the operator, and the degree to which automation is used as intended by the designer. It corresponds to benevolence in Mayer et al. (1995). We follow the model from Lee and See (2004) but divide the three components into more detailed facets for item generation. Three underlying dimensions of trust in automation were postulated: *Reliability/Competence*, *Understandability/Predictability*, and *Intention of Developers*. As already mentioned, trust exhibits a stable individual component (see Section 7.3 for a detailed discussion). Individuals consistently vary in their general propensity to trust, depending on their developmental experiences, personality type, and cultural backgrounds. Additionally, not objective characteristics but a person's subjective perception of a system's characteristics determines trust in automation in the end (Lee & See, 2004; Merritt & Ilgen, 2008). We, therefore, added the individual component, *Propensity to Trust*, from the model of Mayer et al. (1995) as a moderator but also as a direct determinant of trust in automation.

The model of Mayer et al. (1995) addresses interpersonal trust. While other human individuals may be perceived more or less as individuals, different driving automation systems seem to be perceived as a single technology (Schoettle & Sivak, 2014b). This increases the importance of prior familiarity because trust is thereby probably not evaluated again for each driving automation system. Familiarity is assumed to have an indirect influence on trust in automation. With increasing familiarity, operators form expectations, calibrate their trust, and eventually, their confidence in the evaluation of the attributes increases (Hergeth, Lorenz, & Krems, 2017). For example, if no unexpected failures occur, the confidence in the system's reliability increases. As experience with a system grows, trust builds up until a certain level is reached (Beggiato et al., 2015). Taking this into account, *Familiarity* with an automated system was included as a moderator in the theoretical model. Figure 18 illustrates the complete model structure. Based on Mayer et al. (1995), we define trust in automation as the attitude of a user to be willing to be vulnerable to the actions of an automated system based on the expectation that it will perform a particular action important to the user, irrespective of the ability to monitor or to intervene.

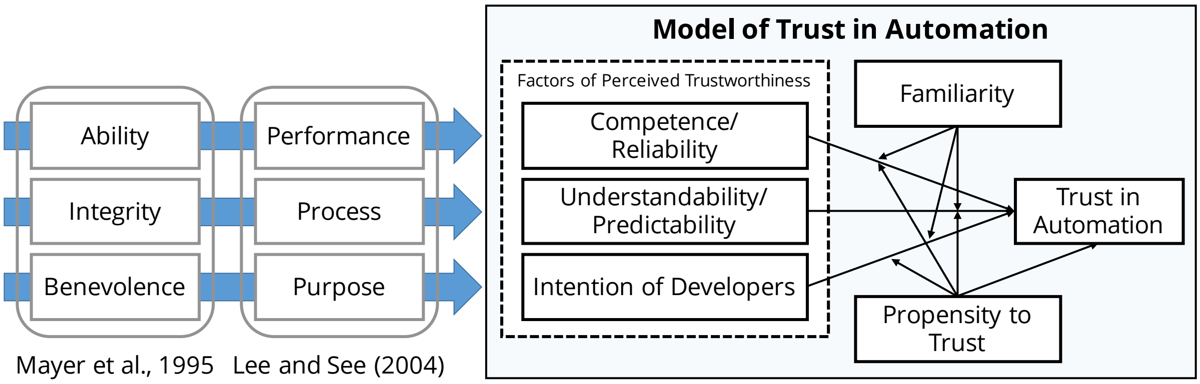

Figure 18. Model of trust in automation based on the postulated dimensions by Mayer et al. (1995) and Lee and See (2004).

Likert-scales are used as means of measurement. Measurement by Likert-scales is based on summative scaling, where respondents use a ranked scale to indicate their agreement with statements. The total score is obtained by summing the ratings of each statement. The goal is to combine the single item responses of an individual to obtain a total score that represents a reliable measurement – multiple Likert-type items form one coherent Likert scale (Hubley & Zumbo, 2013; Uebersax, 2006). The analysis of such a scale assumes a dominance response process, i.e. an individual exhibiting a high degree in the assessed latent trait is assumed to answer positively with high probability (Drasgow, Chernyshenko, & Stark, 2010). A 5-point rating scale ranging between 1 (= strongly disagree) to 5 (= strongly agree) was chosen as the response format. Rating scales with a very fine-grained range, for example from 1 to 100 as in Brown and Galster (2004), offer a resolution that might be inadequate for the provided precision of the measurement, resulting in merely artificial precision. Furthermore, the self-report of trust is based on introspection. It is questionable whether the participants are able to access their trust by introspection with such a granularity as provided by the scale. Measurement, as defined by Krantz, Luce, Suppes, and Tversky (2007), is the process of mapping empirical relational structures onto numerical relational structures. A scale maps the relations of the empirical objects to numerical values. There are different classes of scales, which differ in how much information/relations between the empirical objects is mapped onto numbers. Such a fine-grained scale might map an empirical structure, which does not exist in this resolution, onto numbers with limited meaning. If such scales provide no anchor points, measurement at interval scale level is also even more problematic since equidistance between the rating scale points becomes even more questionable.

We followed a deductive approach for the generation of items (Burisch, 1978) and constructed the questionnaire based on classical test theory (Moosbrugger & Kelava, 2012). An initial set of 57 items was generated. Approximately one third of the items was inversely formulated to reduce response bias (e.g., acquiescence bias) and based on Likert's notion that someone with a positive attitude about the object should also disagree with negative statements. An online survey was conducted for item analysis. The participants watched two videos: 1) an explanation of the underlying technology of automated vehicles (10 min), 2) a conditionally automated highway drive (3 min). A total of $n = 94$ participants completed the survey, 32 participants were female (34.0 %), 60 were male (63.8 %). The mean age was $M = 35.60$ years ($SD = 14.60$, ranging from 17 to 71 years). Based on the criteria item difficulty, standard deviation, item-total correlation, internal consistency, overlap with other items in content, and response quote, 32 items were eliminated, leaving 25 items.

The first validation was carried out in a subsequent online study. In a between-subjects design, a sample of $n = 58$ participants (age range 17 to 72, mean age $M = 34.00$ years, $SD = 15.10$, 58.6 % male, 37.9 % female) watched a video of a conditionally automated highway drive. Participants were randomly assigned to a *reliable* condition, where the video showed a perfectly functioning automation, or a *non-reliable* condition, where participants watched an extended version including a take-over request. As expected, participants of the reliable condition rated the ADS more reliable ($t(41.32) = 3.76$, $p < .001$, $d = 1.05$). Additionally, participants rated their trust directly by answering the item “I trust this system” on a 5-point rating scale ranging between 1 (= strongly disagree) to 5 (= strongly agree). All scales correlated positively with different strength with this rating (lowest: *Familiarity*: $r = .33$; highest: *Reliability*: $r = .85$). Although the total questionnaire correlated strongly with this item ($r = .81$), we found no significant difference between the two conditions ($t(46.92) = 1.21$, $p = .23$, $d = 0.33$), on the contrary for the direct question ($t(45.63) = 2.58$, $p = .01$, $d = 0.71$). Because of their high correlation, the scales *Competence* and *Reliability* were merged, leading to a reduction to 17 items. The internal consistency of the scales ranged from acceptable ($\alpha = .75$; *Propensity to trust*) to excellent ($\alpha = .92$; *Reliability/Competence*).

McCoach et al. (2013) recommend utilizing an exploratory factor analysis (EFA) to evaluate the structure in the very first pilot study because it allows for the highest flexibility of potential solutions. An exploratory factor analysis was conducted to assess whether the structure of the covariation among items is consistent with the proposed factor structure of the trust model. The analysis was performed in JASP (JASP Team, 2018). The dataset showed a sufficient basis to conduct an initial exploratory factor analysis (KMO = .80, Bartlett-Test: $\chi^2(136) = 418.81$, $p < .001$). Following the recommendations of Sakaluk and Short (2017) and McCoach et al. (2013), we chose principal axis factoring as the extraction method and oblique rotation (oblimin) to make the factor solution more interpretable. Parallel analysis by Horn (1965) as well as multiple item factor loadings > .40 on only one single factor determined the extracted factors (Figure 19). Results of the analysis provide initial support for the assumed factorial structure. The resulting pattern matrix (Table 3) shows a clear structure of four factors with high over-determination, “the degree to which each factor is clearly represented by a sufficient number of variables” (MacCallum, Widaman, Zhang, & Hong, 1999, p. 89). Each factor exhibits high pattern coefficients (> .50) by multiple variables while each of the items does not load substantially (> .35) onto other factors, a requirement for a stable solution. Medium to high communalities were observed. Table 4 and Table 5 provide further information on the resulting solution.

Table 3
Pattern matrix generated by principal axis factoring; loadings < .35 have been omitted

	Factor 1	Factor 2	Factor 3	Factor 4	Uniqueness
Familiarity 1			.81		.31
Familiarity 2			.80		.34
Intention of Developers 1		.74			.46
Intention of Developers 2		.49			.45
Propensity to Trust 1				.58	.60
Propensity to Trust 2				.55	.36
Propensity to Trust 3				.59	.55
Reliability/Competence 1	.88				.15
Reliability/Competence 2	.70				.34
Reliability/Competence 3	.79				.23
Reliability/Competence 4	.82				.30
Reliability/Competence 5	.86				.28
Reliability/Competence 6	.70				.44
Understanding/Predictability 1		.65			.36
Understanding/Predictability 2		.60			.44
Understanding/Predictability 3	.64				.24
Understanding/Predictability 4		.62			.50

Table 4
Inter-correlations matrix of the extracted factors

	Factor 1	Factor 2	Factor 3	Factor 4
Factor 1				
Factor 2	.65			
Factor 3	.25	.24		
Factor 4	.19	.31	.04	

Table 5
Fit indices of the resulting model

Chi-squared test			Additional fit indices	
Value	*df*	*p*	RMSEA	TLI
112.139	74	.003	0.09 [0.05, 0.11]	.85

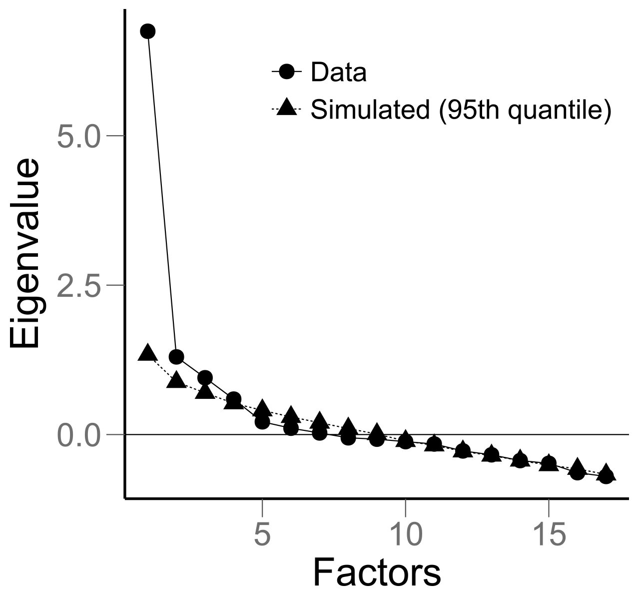

Figure 19. Scree plot of the extracted factors with a parallel analysis by Horn (1965) superimposed.

McNeish (2018) advises against using Cronbach's alpha as a reliability index because its rigid assumptions are routinely violated. He suggests using the omega coefficient, which is conceptually similar to Cronbach's alpha but makes less strict assumptions. In fact, omega total is a more general version of Cronbach's alpha: It also assumes unidimensionality, but the items are allowed to vary in how strongly they are related to the measured construct. Revelle's omega differs from omega total in its more sophisticated variance decomposition. Given that the items each implement a 5-point rating scale, relying on Pearson covariance matrices is reasonable (Rhemtulla, Brosseau-Liard, & Savalei, 2012). All scales exhibited good to excellent internal consistency (Table 6).

Table 6
Indices of the internal consistency of each scale; [a] since Omega total and Revelle's omega cannot be calculated for scales with fewer than three items, the Spearman-Brown coefficient according to Eisinga et al. (2013) was calculated

	Omega Total	Revelle's Omega
Familiarity	.83[a]	-
Intention of Developers	.79[a]	-
Propensity to Trust	.78	.77
Reliability/Competence	.92	.95
Understanding	.81	.88

The factor *Reliability/Competence* was the first extracted factor and, therefore, explained a very major part of the variance, which may be expected given the design of the study, i.e. automation reliability was manipulated between the conditions. However, no factor for *Intention of Developers* could be extracted. The reason for this may lie in the domain of automated driving. A driving automation system is an expensive, highly sophisticated system whose development was motivated by the increase in safety and comfort. The developers of the system are known to be professional car manufacturers. Thus, it is hard to imagine that a driving automation system's developers did not act in a benevolent manner. A revised version of the questionnaire may eliminate this dimension, at least in the domain of automated driving. Item 3 of *Understanding* ("The system state was always clear to me") seems to exhibit a certain degree of multidimensionality and may also be eliminated if this again is the case in future analyses.

Although the aim was to conduct an EFA, fit indices for the model, known from confirmatory factor analysis (CFA), are also reported (Table 5). Fit indices indicate how well the empirical data of the study actually conform to the proposed model. A CFA, therefore, is a more stringent test if the pattern of relationships among the items can be explained by the proposed model/factor structure (McCoach et al., 2013). The chi-squared test evaluates the null hypothesis that the proposed model exactly reproduces the population covariance matrix implied by the data (McCoach et al., 2013). This null hypothesis has to be rejected for the four-factor model. Besides the chi-squared test is generally too liberal at small samples sizes, as in this study, the informative value of this rejection is limited by the fact that a model is always a simplification of a process in reality that never intends to exactly recreate it (McCoach, 2003). The root mean square error of approximation (RMSEA) is an index of absolute fit that compensates for the effect of model complexity (Hu & Bentler, 1999) and can be considered an estimate of the misfit of the model per degree of freedom in the population (Preacher et al., 2013). Cut-offs for small sample sizes ($N \leq 250$) are .08 for a mediocre fit whereas .10 and larger indicates a poor fit (Heene, Hilbert, Draxler, Ziegler, & Bühner, 2011; Hu & Bentler, 1999; MacCallum, Browne, & Sugawara, 1996), indicating a mediocre fit for the four-factor structure of the trust model. However, the estimate is positively biased and the amount of the bias depends on the smallness of the sample and the degrees of freedom (Kenny, Kaniskan, & McCoach, 2014). The Tucker-Lewis Index (TLI) indicates an incremental fit and also compensates for model complexity. A TLI value at or above .95 indicates a good fit, TLI values below .90 are generally considered less than satisfactory (McCoach, 2003). The four-factor model does not fulfill this criterion. However, once again, the TLI is biased in small samples, i.e. it is underestimated in samples with fewer than 100 participants. Heene et al. (2011) echoe previous critique on the application of fixed cut-off rules for model fit because of the multiple dependencies of the fit indices on the conditions (e.g., the achieved factor loadings) and on sample size. After establishing the trust model, two items for measurement of trust in automation itself ("I trust the system" and "I can rely on the system") forming the subscale *Trust in Automation* were added.

The EFA gathered sufficient preliminary evidence of the factor structure and shows that further pursuit of the model is reasonable. Nevertheless, this analysis of construct validity is certainly not sufficient. Firstly, the sample size of $n = 58$ participants results in a case/item ratio of approximately 3:1, which reflects the absolute minimum for a sensible analysis and may be too small to produce a stable solution. However, the minimum required ratio is not constant across studies but rather depends on aspects of the variables and study design (MacCallum et al., 1999). Given a clear factor structure, a high degree of over-determination and high communalities (constantly > .60, as in this study), it is nevertheless possible to reach a stable factor solution even with a sample size smaller than 100 participants (McCoach et al., 2013). Secondly, the participants did not experience driving automation themselves but watched videos of it. Thirdly, the participants only got a short, probably first impression of a driving automation system. This may promote a single-factor structure because the participants might not have had enough insight into the driving automation system to form themselves a detailed, multifaceted impression.

The results of the initial exploratory factor analysis established sufficient initial evidence for the factor structure, affirming that further work is sensible but also needed. Thus, the development process for the questionnaire has certainly not yet come to its end. Future studies have to investigate and ensure the construct validity in greater detail and need to investigate the structure in an applied setting with an adequate sample size. Future work should also follow up this analysis with a CFA to put the established structure to a more rigorous test. In a structural equation model, the claimed paths and relationships of the model can be directly tested and different models can be compared.

The questionnaire's criterion validity was examined in its first use in a driving simulator study in Article 4, where the developed questionnaire to measure trust in automation was used in an applied setting for the first time. In this driving simulator study, 40 participants encountered three critical situations while driving in a conditionally automated vehicle (SAE Level 3) on a highway while being engaged in an NDRT. Eye tracking was used to assess how much the participants rely on the ADS. Furthermore, the instruction for the ADS was varied between two groups with participants receiving either trust-promoting (*Trust promoted* group) or trust-lowering (*Trust lowered* group) introductory information. The trust questionnaire was administered three times: 1) after an introductory video, 2) after an introductory drive, 3) after the experimental drive. It was expected that, firstly, self-reported trust will correlate positively with reliance on automation and, secondly, that participants of the *Trust promoted* group will report higher trust than the *Trust lowered* group.

The analysis comprised the whole *Trust in Automation Questionnaire* (TiA; 19 items) as well as just the subscale *Trust in Automation* and the subscale *Competence*. Regarding the reliability of the *Trust in Automation* subscale, the drawbacks of short scales become eminent. The scale exhibits a low reliability of $\alpha = .63$ after the video and of $\alpha = .70$ after the introductory drive, while it achieved a high reliability of $\alpha = .85$ after the experimental drive. This reflects the problems mentioned earlier with single-item scales: They are more vulnerable to random measurement errors and more susceptible to unknown biases in meaning and interpretation (Emons et al., 2007; Hoeppner et al., 2011). Nevertheless, the subscale *Trust in Automation* was the scale that showed the largest difference ($M_{\text{diff}} = 0.45$, $d = 0.59$, $BF_{-0} = 4.35$) between the two groups after the introductory drive. The subscale might be more sensitive than the whole questionnaire, but this does not guarantee that its predictive performance regarding trust in other systems is superior – predictive quality might vary in different situations and context. The experiment included two situations (Situation 1: overtaking maneuver; Situation 2: adapting speed to a headway vehicle) that were solved by the automated vehicle, but a take-over was a reasonable action if one does not trust automation. In both situations, participants who intervened showed lower trust than participants who did not intervene. The effect size was comparable between the full TiA questionnaire (Situation 1: $d = 0.41$, Situation 2: $d = 0.51$) and the subscale *Trust in Automation* (Situation 1: $d = 0.50$, Situation 2: $d = 0.45$). The same results were obtained for the take-over situation, where participants who crashed reported higher trust than collision-free participants (Full TiA: $d = 0.51$; subscale *Trust in Automation*: $d = 0.58$). Both scales correlated moderately with take-over time (Full TiA: $r = .27$; subscale *Trust in Automation*: $r = .33$) and minimum TTC (Full TiA: $r = -.29$; subscale *Trust in Automation*: $r = -.35$). Both full questionnaire and subscale *Trust in Automation* correlated with the participants' gaze behavior with the expected sign and at approximately the same magnitude (medium effect) in all three measurement intervals.

In summary, participants with higher trust scores consistently showed stronger reliance in all behavioral measurements compared to participants with a lower trust score. Consequently, the study confirms the predictive validity of the questionnaire. Furthermore, the medium-sized correlation between the TiA questionnaire score and the affinity for technology questionnaire (Feuerberg, Bahner, & Manzey, 2005) of $r = .47$ (BF = 18.85) shows that trust is related to affinity for technology, yet it represents a distinct construct, supporting its construct validity.

The two-item subscale *Trust in Automation* showed lower reliability but was more sensitive regarding group differences and performed equally as well as the full TiA questionnaire regarding all other measures. This provokes the question of whether a single-item scale may be sufficient for a valid measurement of trust. The benefits of using single-item measures have been listed by several researchers (Fuchs & Diamantopoulos, 2009; Hoeppner et al., 2011): Single-item scales are less monotonous and time-consuming. They can also be administered during an experiment for a momentary assessment, for example while driving. The aforementioned advantages of multi-item scales are also accompanied by drawbacks, such as boredom caused by redundant items and fatigue in lengthy questionnaires (Burisch, 1984). Nevertheless, for a detailed assessment of a multidimensional construct such as trust in automation, a multi-item measure is typically necessary (Nunnally & Bernstein, 1994).

Yet, Fuchs and Diamantopoulos (2009) argue that the use of a single-item scale may still be appropriate in certain cases. For example, Sloan, Aaronson, Cappelleri, Fairclough, and Varricchio (2002), while discussing the quality of life measurement, claim that "there comes a point where the construct becomes so complex that a single question may be the best approach" (p. 481). Hence, when measuring overall job satisfaction, the best measurement may be a question like "Overall, how satisfied are you with your job?" (Fuchs & Diamantopoulos, 2009, p. 204; Scarpello & Campbell, 1983). A single item on trust in automation reflects the conceptualization of trust in automation as a mainly affective response with influences from analytic and analogical processes. Lee and See (2004) suggest that because of the complexity of automation technology, operators probably rely less on analytic calculations to guide their behavior but rather apply heuristics to accommodate the limits of the human bounded rationality (Gigerenzer & Selten, 2002). A situation might occur where operators cannot form a complete mental model of an automated system as it is too complex to perfectly predict its behavior. Emotions can then guide behavior when rules are not effective or when cognitive resources are too limited for a calculated rational choice (Damasio, 1996; Lee & See, 2004).

In the validation study, 78 % of the participants have had no contact with conditionally automated driving before. Thus, it might not have been possible for the participants to rate each dimension of the trust questionnaire adequately because of a lack of knowledge or experience.

Differences in the ability to accurately rate a system have been pointed out by Annett (2002) who gives the example of expert test drivers who learn by experience to identify and rate the subtle dynamic features of a vehicle. It is conceivable that the participants' trust rating was a rather global impression or rating, which can be captured accurately by a single item. It is unclear if participants would also provide a global rating if they had more experience with an automated vehicle.

Yet, such a simplification of the construct trust in automation comes with a cost: The *Trust in Automation* scale consists of two items, one of them with the content "I can rely on the system". It is not surprising that such a measure highly correlates with behavioral reliance measures such as eye tracking and intervention frequency. For such a narrow conceptualization of trust, the high validity may justify the use of a single-item measure (Flake et al., 2017). The construct trust in automation, which possesses a detailed underlying theory (Lee & See, 2004; Mayer et al., 1995), would then, at the same time, become one with its measure and loses any theoretical meaning beyond that measure (Bagozzi, 1982). This measurement would then be in conflict with the definition of what it intends to measure. Indeed, as already mentioned, trust is an attitude that stands between the belief about characteristics of an automated system and the intention to rely on it. Attitude, intention, and actual behavior are not in a deterministic but in a probabilistic relationship (Ajzen & Fishbein, 1980). Whether trust translates into actual reliance on an automated system is also influenced by other factors such as self-confidence or time constraints (Dzindolet et al., 2001; Lee & See, 2004; Meyer, 2004).

The use of a single-item measure is also problematic in longitudinal studies: If the observed value changes, it is not possible to differentiate between a true change in the construct and a change caused by imperfect reliability of the measurement (Fuchs & Diamantopoulos, 2009). Here, researchers may fall back on the multi-item questionnaire. If a single-item is administered to obtain a global assessment, it has to be taken into account that the respondents each consider an individual set of aspects of trust and of the automated system, weighted by their own individual preferences, providing a tailor-made impression (Nagy, 2002). Hence, respondents may not consider the same aspects or may not even think of a relevant aspect at all. It, thereby, remains unknown how the assessment is constituted. To ensure that each participant assesses the same construct, i.e. that a common understanding of trust exists, an accurate definition of trust in automation has to be provided in this case (Fuchs & Diamantopoulos, 2009). On the other hand, multi-item scales are less individual but more comparable. A preset of aspects, formed by the questionnaire's scales, also helps and guides the participants to rate the system. Multiple scales also provide the possibility to express the trust rating in greater detail. With a single-item scale, should the trust score turn out to be low, the researchers then have no indication for the reason. Contrarily, multiple scales may enable researchers to find the cause in a certain characteristic of the automated system. For example, it could be perceived as reliable, but participants did not understand its functioning. Thus, it is reasonable to use a multiple-item scale such as the TiA if the aim is a thorough, multi-faceted assessment.

In conclusion, if the research objective is a global assessment, an overall feeling or impression by the participants, then a single-item may provide all the desired information (Fuchs & Diamantopoulos, 2009). It represents a useful supplement that might be sufficient for a single and quick, yet valid assessment and "can provide an acceptable balance between practical needs and psychometric concerns" (Robins et al., 2001, p. 152). This is particularly true if trust is merely used as a moderator or as a control variable (Fuchs & Diamantopoulos, 2009). If the goal is a detailed assessment of trust in automation or if a longitudinal design is implemented, then the multi-item questionnaire may be preferred.

The questionnaire's further development certainly needs to address its psychometric qualities. The low internal consistency of the subscale *Trust in Automation* at the beginning of the study raises the question of whether a short scale of two rather direct items is sufficient as a measurement of trust itself. Mayer and Davis (1999) provide a questionnaire for their model of interpersonal trust, which includes a four-item scale to assess trust. The items are less direct than the two *Trust in Automation* items of the TiA questionnaire and rather aim at the willingness to be vulnerable, corresponding to their definition of trust (Mayer et al., 1995). Thus, a revised version of the TiA questionnaire may adopt this approach and offer a four-item scale (besides the original scales) for trust in automation that is closer to its definition by Körber, Baseler et al. (2018). Items from Mayer and Davis (1999) adapted to the domain of automation could, for example, read "I would be comfortable handing over the driving task to the driving automation system without monitoring it" or "If I had my way, I wouldn't let a driving automation system have any influence on the driving task". A single item for assessing trust in automation such as "I trust this driving automation system" then may function as the aforementioned pragmatic variant alongside the multi-item questionnaire. In addition, information on the questionnaire's discriminant validity is still missing. Also, further data on the questionnaire's predictive performance have to be gathered.

A revision may also reconsider the inclusion of the scale *Familiarity*. Familiarity itself is not an element of trust in automation but indirectly influences it as a moderator. With increasing familiarity, operators form expectations and the confidence in their evaluation of the attributes increases. If this moderating role is of no interest in a study, the scale could be eliminated to shorten the questionnaire. A core questionnaire only containing the factors that directly influence trust then may be more appropriate. Beyond this, familiarity could also induce response bias: Low familiarity with an automated system could induce a tendency towards a global evaluation of the system due to a lack of in-depth knowledge. It would, therefore, be interesting to administer the questionnaire to participants who are already very familiar with a driving automation system. This is especially of interest regarding the difference between the predictive performance of a single-item measure and the multi-item TiA questionnaire.

In closing, it has to be considered that the measurement of trust in automation by means of a questionnaire certainly has to be viewed in perspective of its position in measurement theory. There have been concerns doubting the possibility of measurement of psychological constructs and their quantitative nature in general (Michell, 1997). However, using rating scales for the measurement of psychological constructs, such as trust, does not exclusively have to be regarded as a form of measurement in the strict sense of the term, i.e. in terms of the representational theory of measurement, where a homomorphic representation of physical empirical relations is mapped to numerical relations (Annett, 2002; Krantz et al., 2007). Instead, following a model-based account of measurement, measurement of trust can rely on an abstract model that is valid for the prediction of an individual's performance during a certain task (Tal, 2017). As Tal (2017) argues, such a model is defined by theoretical and statistical assumptions about the measured psychological construct and its relation to the measurement task. Inference from the indication of a measurement instrument (e.g., a rating scale) to the measurement outcome is non-trivially derived from the model. Measurement is then the coherent and consistent assignment of values to parameters in this model, based on instrument indications. The model defines the content of the measurement outcome, which does not have to hold a counterpart in the observable world – a construct, in the end, is a concept, model, or schematic idea (McCoach et al., 2013). As for the measurement of intelligence, the values do not represent physical properties but empirical relationships between theoretical constructs and other constructs or behavior (Annett, 2002). Trust measurement, thus, may not deliver meaningful, absolute values per se, but values that are meaningful in the context of a model of trust, which is defined by theoretical and statistical assumptions such as confirmed construct validity. In this way, the measurement outcome can be used to predict and explain behavior, decisions, or performance. For this reason, it is unreasonable to apply the same standards to the measurement of trust as to measurements such as take-over time. Nevertheless, the results in Article 4 show that the questionnaire produces meaningful measures with relation to observable and safety-relevant behavior.

B Article 1: Potential individual differences regarding automation effects in automated driving

Körber, M., & Bengler, K. (2014). Potential individual differences regarding automation effects in automated driving. In C. S. G. González, C. C. Ordóñez, & H. Fardoun (Eds.), *Interacción 2014: Proceedings of the XV International Conference on Human Computer Interaction* (pp. 152–158). New York, NY, USA: ACM.

C Article 2: The influence of age on the take-over of vehicle control in highly automated driving

Körber, M., Gold, C., Lechner, D., & Bengler, K. (2016). The influence of age on the take-over of vehicle control in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *39*, 19–32.

D Article 3: Prediction of take-over time in highly automated driving by two psychometric tests

Körber, M., Weißgerber, T., Kalb, L., Blaschke, C., & Farid, M. (2015). Prediction of take-over time in highly automated driving by two psychometric tests. *Dyna*, *82*(193), 195–201.

E Article 4: Introduction matters: Manipulating trust in automation and reliance in automated driving

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, *66*, 18–31.

F Article 5: Vigilance, boredom proneness and detection time of a malfunction in partially automated driving

Körber, M., Schneider, W., & Zimmermann, M. (2015). Vigilance, boredom proneness and detection time of a malfunction in partially automated driving. In *International Conference on Collaboration Technologies and Systems (CTS)* (pp. 70–76). IEEE.