Trust in highly automated driving

vorgelegt von

M. Sc.
Amelie Stephan
ORCID: 0000-0003-3160-9639

von der Fakultät V – Verkehrs- und Maschinensysteme
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktorin der Naturwissenschaften
– Dr. rer. nat. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. Manfred Thüring
Gutachter:	Prof. Dr. Dietrich Manzey
Gutachter:	Prof. Dr. Mark Vollrath

Tag der wissenschaftlichen Aussprache: 21.03.2019

Berlin 2019

Disclaimer

Ergebnisse, Meinungen und Schlüsse dieser Dissertation sind nicht notwendigerweise die der Volkswagen AG.

The results, opinions, and conclusion of this thesis are not necessarily those of Volkswagen Aktiengesellschaft.

Danksagung

Diese Dissertation entstand an der Technischen Universität Berlin in Kooperation mit der Konzernforschung der Volkswagen Aktiengesellschaft und dem Electronics Research Laboratory der Volkswagen Group of America. Ich möchte mich auf diesem Wege bei allen Personen bedanken, die mich während der Bearbeitung meiner Dissertation unterstützt und damit zum Gelingen dieser Arbeit beigetragen haben.

An erster Stelle danke ich meinem betreuenden Doktorvater, Prof. Dr. Dietrich Manzey, der mir in diversen Treffen und Telefonaten stets mit Rat und Tat zur Seite stand und ohne den diese Arbeit nicht zustande gekommen wäre. Herzlichen Dank dafür. Prof. Dr. Mark Vollrath gilt mein Dank für die Übernahme der Zweitkorrektur. Weiterhin danke ich den wissenschaftlichen Mitarbeitern des Institutes für Arbeits- und Organisationspsychologie für ihr wertvolles Feedback und den hilfreichen Austausch im Kolloquium.

Besonders danken möchte meinen Betreuern im Unternehmen, die mich während der gesamten Zeit immer wieder motivierten und deren ehrliche Kritik, Anregungen und Denkanstöße mir stets ein Quell der Inspiration waren. Ihre Anmerkungen haben maßgeblich zur Qualität dieser Arbeit beigetragen, und ihre organisatorische Unterstützung hat mir bei der Fertigstellung dieser Arbeit sehr geholfen. Ich danke allen Kollegen, Doktoranden, Praktikanten und geduldigen Korrekturlesern für die Unterstützung und Kollegialität zu jeder Zeit. Sie alle haben zu einem hervorragenden Arbeitsklima beigetragen und die gemeinsamen Diskussionen haben diese Arbeit immens bereichert. Auch für die technische Unterstützung vieler Kollegen bei der Umsetzung von Nutzerstudien im prototypischen Fahrzeug und im Fahrsimulator sei an dieser Stelle herzlich gedankt – ohne ihre Hilfe wäre die Durchführung der Arbeit in dieser Form nicht möglich gewesen.

Von ganzem Herzen möchte ich insbesondere meiner Familie und meinen Freunden für die moralische Unterstützung, die Kraft und den Rückhalt auch in schwierigen Zeiten danken.

Veröffentlichungen

Konferenzbeiträge

Stephan, A. (2015). Trust in automation – HMI solutions for piloted driving. *2nd International VDI Conference Automated Driving*, Frankfurt, Germany.

Bendewald, L., Glaser, E., Petermann-Stock, I., and Stephan, A. (2015). „Jack" – A holistic approach of designing a human machine interface for highly-automated driving. In *VDI-Berichte – International Congress Electronics in Vehicles* (Vol. 17, pp. 453—467). Düsseldorf: VDI-Verlag GmbH.

Bauerfeind, K., Stephan, A., Hartwich, F., Othersen, I., Hinzmann, S., Bendewald, L. (2017). Analysis of potentials of an HMI-concept concerning conditional automated driving for system-inexperienced vs. system-experienced users. In De Waard, Dick; Di Nocera, F.; Coelho, D.; Edworthy, J.; Brookhuis, Karel A.; Ferlazzo, F.; Franke, T.; Toffetti, A. (Hrsg.). *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2017 Annual Conference*.

Poster

Stephan, A. (2017). *Trust in Automated Driving – The Influence of System Characteristics on Trust in an Automated Vehicle*. Poster presented at the Human Factors and Ergonomics Society Europe Chapter 2016 Annual Conference.

Zusammenfassung

Die Automobilindustrie steht an der Schwelle zu einer neuartigen Technologie: selbstfahrende Fahrzeuge. Solche hochautomatisiert fahrenden Fahrzeuge sind technisch immer besser realisierbar, und Konzerne und Forschungsinstitute auf der ganzen Welt investieren Zeit und Geld, um die einst futuristische Vision auf die Straße zu bringen. Die Technologie wird mit dem Ziel entwickelt, dem Fahrer die manuelle Fahrzeugsteuerung abzunehmen. Dadurch soll sie den Fahrkomfort erhöhen und vor allem zur Verbesserung der allgemeinen Verkehrssicherheit beitragen.

Über die weitere technische Entwicklung hinaus werden psychologische Aspekte und die Gestaltung eines optimalen Nutzererlebens bei der Betrachtung hochautomatisierter Fahrfunktionen immer wichtiger. Insbesondere muss für eine zukünftige gesellschaftliche Nutzung zunächst das Vertrauen in diese Art der Fahrfunktionen aufgebaut werden. Andernfalls, wenn die Menschen nicht bereit sind die Kontrolle einem solchen Fahrzeug anzuvertrauen, wird es nicht genutzt und das Potenzial des hochautomatisierten Fahrens kann nicht voll ausgeschöpft werden.

Das Ziel dieser Arbeit ist es, einflussreiche Faktoren hinsichtlich des Vertrauens in hochautomatisiert fahrende Fahrzeuge festzustellen und zu prüfen, wie dieses Vertrauen durch ein spezifisches Human-Machine Interface (HMI) unterstützt werden kann. Zu diesem Zweck wurden drei Haupt-Untersuchungen mit Probanden durchgeführt. Verschiedene HMI-Konzepte wurden in diesen Nutzerstudien sowohl in einem prototypischen Fahrzeug auf öffentlicher Straße als auch im Fahrsimulator getestet. Ziel der ersten Realfahrtuntersuchung ($N = 28$) mit dem hochautomatisiert fahrenden Fahrzeug war es, einflussreiche Faktoren für das Vertrauen in ein solches Fahrzeug zu untersuchen. Als relevante Faktoren wurden das Persönlichkeitsmerkmal Kontrollbedürfnis sowie eine allgemeine Einstellung gegenüber Technik identifiziert. Die wichtigste Rolle für das Vertrauen spielte jedoch die wahrgenommene Fahrleistung des Systems. In der zweiten Nutzerstudie ($N = 72$) wurde mithilfe einer simulierten Umgebung der Einfluss von Systemgrenzen auf das Vertrauen überprüft. Es konnte nachgewiesen werden, dass die Art der erlebten Systemgrenze eine entscheidende Rolle spielt. Vor allem die Nicht-Detektion eines relevanten Ereignisses in der Fahrsituation minderte das Vertrauen, während eine fälschliche Detektion kaum zu einer Vertrauenssenkung führte. Über mehrere Versuchstage hinweg wurde in einer dritte Nutzerstudie ($N = 18$) untersucht, wie sich das Vertrauen über einen Erstkontakt mit einem hochautomatisiert fahrenden Fahrzeug hinaus entwickelt. In dieser Realfahrtstudie zeigten sich erste Hinweise darauf, dass die Relevanz des HMIs im Verlauf der Systemnutzung zunimmt.

Ein anhand von bisherigen Erkenntnissen und Theorien aufgestelltes Vertrauensmodell wurde mit Hilfe dieser Studien auf den neuen Kontext des hochautomatisierten Fahrens übertragen. Weiterhin wurden Empfehlungen zum Design eines HMI-Konzepts für hochautomatisierte Fahr-

zeuge zusammengetragen und angewendet. Damit unterstützen die Erkenntnisse dieser Arbeit Entwickler bei der Gestaltung von HMI-Konzepten zur Förderung des Vertrauens in automatisierte Fahrfunktionen. Auch wenn der Fahrer zukünftig möglicherweise keine Fahraufgaben mehr übernehmen muss, wird empfohlen ihm zur Unterstützung des Vertrauensaufbaus ein adäquates HMI-Konzept zur Verfügung zu stellen.

Abstract

The automotive industry is on the verge of a new technology: self-driving vehicles. Such highly automated driving vehicles are more and more technically feasible, and corporations and research institutes all over the world are investing time and money to bring the once futuristic vision on the road. The technology is developed with the goal to release the driver from the manual task of controlling the vehicle. Through that, it shall increase driving comfort and, above all, contribute to the enhancement of overall road safety.

Beyond further technical development, psychological aspects and the creation of an optimal user experience gain importance for highly automated driving functionality. In particular, trust in this kind of functionality has yet to be built up for future societal usage. Otherwise, if people are not willing to entrust control to such a vehicle, it will not be used and the potential of highly automated driving cannot be fully exploited.

The aim of this work is to identify influential factors on trust in highly automated driving vehicles and to examine how this trust can be supported by a specific human-machine interface (HMI). To this end, three main studies were conducted with participants. Different HMI concepts were tested in these user studies in a prototype vehicle on public roads as well as in a simulated environment. The aim of the first real-driving study (N = 28) with the highly automated driving vehicle was to test influential factors on trust in such a vehicle. The personality characteristic desire for control as well as a general attitude towards technology were identified as relevant factors. However, most important for trust was the perceived performance of the system. In the second user study (N = 72), the influence of system boundaries on trust was examined with the help of a simulated environment. It was proven that the type of the experienced system limit plays a crucial role. In particular, the non-detection of a relevant event within the driving situation diminished trust, while a false detection led to little trust reduction. Over several trial days, it was examined in a third user study (N = 18) how trust develops beyond a first contact with a highly automated driving system. In this real-driving study, first indications were found that the relevance of the HMI increases with prolonged system use.

A trust model set up based on previous insights and theories was transferred to the new context of highly automated driving with the help of these studies. Furthermore, guidelines for the design of an HMI concept for highly automated vehicles were collected and applied. Thereby, the insights of this work support developers in designing HMI concepts to promote trust in automated driving functionality. Even if the future driver no longer needs to take over driving tasks, it is recommended to provide an adequate HMI concept supporting trust development.

Contents

1 Introduction

A self-driving vehicle is one of the most desirable visions in modern transport (Maurer, Gerdes, Lenz, & Winner, 2015). The new technology of automated driving aims at increasing road safety and the reduction of road fatalities, while at the same time it is supposed to promote driving comfort and convenience significantly. The potential benefits for drivers are countless: their travel time could be used more productively by working or simply relaxing in the car during their daily commute. Moreover, this time could be reduced because of designated lanes and improved traffic fluency. A fully automated vehicle could even pick its owner up at home, or find a parking spot on its own. Already in the year 1956, an advertisement for America's Independent Electric Light and Power Companies by Miller demonstrated the possible advantages of automated driving, illustrating a family enjoying a ride in their autonomous vehicle. The advertising text said "Electricity may be the driver. One day your car may speed along an electric super-highway, its speed and steering automatically controlled by electronic devices embedded in the road. Highways will be made safe—by electricity! No traffic jams... no collisions... no driver fatigue." (America's Independent Electric Light and Power Companies, 1956, p. 8).

Modern technology renders possible what has been science fiction half a century ago. Today, automated vehicles are not far-fetched any longer, but actually achievable and already on the road at some places. Already in 2010, the technology company Google announced that their self-driving cars had accomplished 140,000 miles of automated driving, and in 2016 the company reported over 600,000 miles. Several Original Equipment Manufacturers (OEMs) announced automated driving capabilities of some sort in series production by 2020 (Toyota Motor Corporation, BMW AG, Daimler AG, Nissan, Volvo Car Corporation, Audi AG, Tesla, etc.). Other companies are entering the competition as well, one example being the online transportation network company Uber that started a partnership with Carnegie Mellon University in 2015 to develop autonomous cars (Harris, 2015). The race is on for automotive and technology companies to prove their expertise and progressiveness in this area, as advanced driving assistance and even automated driving become more of a requirement rather than a gimmick for contemporary vehicles.

1.1 Motivation

A challenge that is becoming more and more important for the actual launch of automated driving is the formation of trust of potential users in such vehicles. Although prototype vehicles are present, the future users never had the chance to give the new technology a try until now. Drivers might not necessarily want to hand over control to the vehicle without knowing how the system will react in various kinds of driving situations. Even though human error is involved in over 90% of all traffic accidents (National Highway Traffic Safety Administration [NHTSA], 2015), not all drivers feel the need to be assisted. 90% rather feel they belong to the better half of all drivers (Svenson, 1981). A recent study of Eimler and Geisler (2015) has even shown that 70% are of the opinion that humans are better drivers than automated vehicles. In addition, when asked if they like to drive themselves or if they like to be passengers, most people want to be drivers—partly because they like the fun of driving, and partly because they want to be in control of the situation. From a recent compilation of surveys in Germany follows that 27% of all drivers are unwilling to hand control over to the vehicle in any situation (Statista, 2015). 63% are willing to let the car park, and 45% would use an automated system during traffic on highways. According to the survey, only 7% would dare to hand control to the vehicle during a complete drive. In another large international survey, fully automated driving has been found to be a fascinating idea, however the survey still identified manual driving as the most enjoyable driving mode (Kyriakidis, Happee, & De Winter, 2015).

Indeed, the development of more and more automated functionality in series vehicles brings with it some issues (so-called *automation effects*) formerly only known from domains like the industrial sector and aviation. These drawbacks of automation have to be acknowledged in the automotive area as well. Once people are released from the manual execution of the driving task itself, drivers might not pay attention to the driving situation when it actually might be necessary. Phenomena like decreasing situation and mode awareness during automation use are likely to arise and need to be countered. Without continuous manual activity and involvement, people might even lose their driving skills and might not be able to intervene manually anymore after longer periods of using automated driving technology.

Human factors aspects of automated driving technologies are of major significance for a successful launch of automated driving vehicles (Walker, Stanton, & Young, 2001). Trust is of paramount importance in the development of all autonomous systems designed to support and relieve humans. Trust will not only affect the willingness to purchase an automated driving system, but will as well influence the extent to which drivers agree to let the car drive them and actually spend their time relaxing rather than being anxious while the vehicle is in control. When designing the optimum user experience for drivers in an automated vehicle, it is particularly relevant to secure that trust in the system can be established, so that the driver is comfortable in relinquishing the driving task and transferring it to the system. The references mentioned above show that trust in automated driving systems should not be taken for granted. Today's drivers need to be introduced to the new technology with care, and need to be supported to overcome their hesitant reserve.

To make sure the new technology is as safe and comfortable as envisaged, numerous topics need to be addressed. The new role of the driver needs to be clarified. It is necessary to provide a strategy to generate trust in self-driving vehicles to exploit the full capability of the technology and gain the most value out of it. Finally, it needs to be determined what awareness and capabilities the driver needs to maintain when he has abandoned his role of being the main operator. Technically, when the vehicle takes over control, it is in charge of every maneuver and every reaction to the surroundings (depending on the specific level of automation). In fact, the driver does not need to monitor or supervise the car until it indicates that the driver shall take over control again because system boundaries are reached or a similar reason—as long as the driver sufficiently trusts the system. The behavior of the vehicle and its driving performance will be of major importance in this process. However, conveying information about the driving system and the environment to the driver is crucial as well. On the one hand, it is necessary to ensure awareness of the current mode and situational awareness in case the driver needs to take over driving. Whenever the system is not capable of handling a situation, it should indicate this to the driver and hand over the control of the vehicle. At this moment, it needs to be made sure that the driver is able to get back in the loop during the takeover process. The driver needs to understand the situation around him as well as the system's intentions. On the other hand, an accurate human-machine interface (HMI) concept might also help to strengthen trust and confidence in the system by explaining the system's behavior. By enhancing the driver's understanding of the automation, it could foster trust in an automated driving system. Therefore, when designing future automated driving systems, the goal should be to release the driver of the (at times) annoying execution of manual actions, but simultaneously provide driving-relevant information in order to make sure the driver can assess the system and the situation at any time (Buld et al., 2002). The design of the interaction should enhance the confidence of the driver during mode changes and provide all information in a way the human operator can easily understand and react to. While these efforts have already been made in aviation automation, they are still in an early phase of their development for automated driving vehicles.

Research on trust in automation has been carried out in diverse domains, since automated systems are widely spread in our everyday lives. Plenty of research on trust in automation is done in the fields of aviation, telerobotics, and production plants, for example, to ensure effective collaboration between man and machine. Numerous studies have also addressed trust in automated driving and provided first insights in this issue (e.g., Gold, Körber, Hohenberger, Lechner, & Bengler, 2015; Helldin, Falkman, Riveiro, & Davidsson, 2013; Hergeth, Lorenz, Krems, & Toenert, 2015). However, due to the novelty of the technology, the studies were thus far conducted in simulated environments, i.e. static or dynamic driving simulators, which might limit their external validity. Field studies are needed to broaden the insights and to validate the findings attained until now. Also, there is a lack of insights regarding long-term development of trust in automated driving systems. It is unclear what consequences might be contained in long-term use of an automated driving vehicle and how the user's trust will develop over time. Once the new technology is ready for the market, customers cannot be left alone with the exploration of it—they need to be carefully made familiar with it in order to trust it. To make the automated driving technology acceptable for users, more research needs to be conducted on the interaction design of these future self-driving cars.

In summary, drivers need to trust the actions of the automated driving system and be willing to use it in order to accomplish the goals of automated driving to increase safety and enhance the comfort of driving. The key question of this work is thus: How can trust in an automated driving system be supported by an HMI concept? This question is approached by studying three main aspects: relevant factors influencing trust, the influence of system performance and HMI on trust, and the development of trust over several system encounters.

1.2 Contribution

The aim of this work is to find out how drivers can be supported by an HMI concept to develop trust in a highly automated driving (HAD) system.

In this work, substantial characteristics of the system as well as human predispositions relevant for the development of system trust are identified for the specific context of automated driving. An HMI concept for automated driving systems is realized and evaluated in a prototype vehicle to identify information relevant for the driver. Empirical evaluations are conducted to specify how system states and information can be conveyed so that the driver develops trust in the automated driving system. To the author's knowledge, no naturalistic driving studies on trust in automated driving had been conducted before the start of this research. This contribution differs from earlier approaches by investigating the topic of trust in automated driving not only in a restrained simulated environment, but also in a naturalistic environment with an automated vehicle under real traffic conditions. Per definition, trust is important in situations of uncertainty and vulnerability. Therefore, for effective research it is essential to create study settings as close as possible to the real situations. Prototype vehicles with automated driving functionality are still rare, and only a limited number of people have access to them. In the user studies presented in this work, drivers have the possibility to discover how it feels to abandon the driving task completely for a certain time. In the initial study, system and human factors that can potentially influence trust in automated driving are analyzed. It is assessed what information drivers need to feel comfortable and be able to hand over control during a real automated drive. Secondly, the effect of system boundaries on trust calibration is investigated in a simulator study with varying system reliability and HMI concepts of differing transparency. Lastly, a longer-term driving study with automated driving functionality is conducted. The technology is still new, and no knowledge about medium- or long-term development of trust in the automated driving technology exists. Within the scope of this thesis, an observation of trust development across multiple practical experiences is made for the first time to investigate how trust develops over time and how the need for information changes over the course of system use. Further, in all studies subjective and objective methods of assessing trust in an automated driving system are considered.

The key point of concern of this work is the investigation of trust in an automated driving system and of ways to support this trust with an HMI concept. It focuses on conditional and high automated driving levels, and does not include semi-automated or fully automated driving (SAE International, 2014). To investigate the topic of trust in automated driving, a driving simulator as well as a prototype vehicle with automated driving technology were used. The vehicle is capable of driving on highways with normal traffic including lane change maneuvers, however it is of course still in a research state and is not completely production-ready yet. Investigations were exclusively made with this prototype vehicle, confining especially the longer-term evaluation to a limited time period of a few weeks in total. Other aspects of automated driving, such as technical, legal, or ethical questions on that topic are not part of the discussion on hand and are merely touched occasionally.

1.3 Overview of this work

The work at hand is structured as follows. Giving an overview of the current status of research, Chapter 2 starts by defining what is understood by automation and automated driving systems. It advances by giving a brief summary of research done on the topic of trust in automation, including recent findings regarding relevant factors and models of trust. The chapter places special emphasis on research regarding the design of system interaction concepts facing the challenges mentioned above. Subsequently, to advance the research already done on that topic, Chapter 3 describes a theoretical working model for trust in the context of HAD. Open research questions are prescribed based on the findings obtained so far, and an accordant HMI concept for HAD is developed based on findings of exploratory studies. Chapter 4 specifies the empirical investigations conducted to address the hypotheses arising from the presented model and the derived research questions. To support the theoretical model with empirical findings, data was obtained to reveal important aspects of trust in an automated vehicle in varying environments. Especially the studies under realistic driving conditions shall help explore what components are truly necessary to enhance trust in the new technology. Chapter 5 summarizes and interprets the conducted research by subsuming all studies' outcomes. Also, the chapter discusses shortcomings of the studies and summarizes the implications of the results for further research.

The work at hand specifically aims at investigating the topic of trust in HAD vehicles. Within the framework of this thesis, determinants and correlates of trust between humans and automated vehicles are explored and ways to engender trust in such a system are identified. The work considers various models and theories of trust derived from other domains, and proposes metrics for measuring trust in this field of research.

Numerous challenges need to be faced to bring automated driving on the road for everyone. In return, automated driving has the potential of changing our mobility dramatically. Future users should be guided into this new world of automated driving.

2 Theoretical background

To investigate the field of trust in an automated vehicle in detail, this chapter first describes the recent developments in automated driving technology (Section 2.1). The definition of automated driving used in this work is provided in Section 2.1.1. Section 2.1.2 goes into more detail on side effects of automation like supervisory control, out-of-the-loop performance, and ironies of automation, followed by an overview of technical and societal conditions that need to be fulfilled to facilitate the implementation of automated driving (Section 2.1.3). In a second step, the state of the art of research on trust in automated driving is outlined in Section 2.2. To establish a joint understanding of the concept, the definition of trust utilized in this contribution is depicted in Section 2.2.1. Related psychological constructs and effects of trust are subject of discussion in Section 2.2.2. Subsequently, research results are summarized in Section 2.2.3, where different human and system characteristics relevant for trust development are explained in detail. Section 2.2.4 gives a review of current trust models brought about by research on this topic. Section 2.3 specifically reports on findings relevant for supporting trust in automated driving. The state of research is summed up in Section 2.3.1, and a collection of relevant design recommendations is presented in Section 2.3.2. Section 2.4 finally summarizes the chapter and derives open research issues.

2.1 Automated driving

Due to enormous progress in technical science and research, humans today can benefit from the potential of modern machines and systems. Regarding vehicles, this does not only mean higher speeds or greater ranges—these days, this also can imply artificial intelligence to a certain degree. Rapid advancements in hard- and software result in new sophisticated driver information and driver assistance systems. Even more, vehicles like modern airplanes are already able to operate completely autonomous for certain periods of time. The automotive industry follows suit, trying to enhance customer benefit by making driving as comfortable as possible.

This work specifically addresses automated driving systems as a special form of automation. A technology that has the capability to drive a vehicle without the active physical control or monitoring by a human driver is considered as automated driving technology (per definition of California Legislative Counsel, 2012).

Automotive manufacturers like AUDI AG, BMW AG, or Daimler AG as well as numerous research institutions collaborate in projects that were established to advance automated driving (Maurer et al., 2015). The research project "Highly automated vehicles for intelligent transport" (HAVEit), for example, aimed at the realization of steps towards HAD and was funded by the European Union (EU) with 17 million Euro (www.haveit-eu.org). Another EU-project is the project "Automated driving applications and technologies for intelligent vehicles" (AdaptIVe), a successor to the project InteractIVe. The consortium of 29 partners aims at developing automated driving functions for different complex traffic situations and driver states and concurrently addresses legal issues that need to be solved for bringing such systems to the market (www.adaptive-ip.eu). The combined research project "Ko-HAF: Cooperative Highly Automated Driving" started in June 2015 and is part of the program "New Vehicle and System Technologies" by Germany's Federal Ministry for Economic Affairs and Energy (BMWi). It specifically aims at higher levels of automated driving and the vehicle's communication with both the driver and other highly automated vehicles. One main objective of all those endeavors is the *vision zero*, a vision of accident-free traffic and zero road fatalities (Maurer et al., 2015). Not only in the industry are research projects pursued with enthusiasm. A breakthrough in work on vehicles driving autonomously constituted the Grand Challenge by the Defense Advanced Research Projects Agency (DARPA). As one of the greatest events for autonomous driving, it took place in the years 2004 and 2005 and was extended to other events like the DARPA Urban Challenge in 2007. It motivated organizations as well as universities from around the world to compete on a test track with their fully autonomous ground vehicles and to prove their technology is capable of completing an off-road course through the desert (or a city course in 2007) within a limited time (Thrun et al., 2006; Urmson et al., 2008).

2.1.1 Definition of automation

Parasuraman and Riley (1997, p. 231) define automation "as the execution by a machine agent (usually a computer) of a function that was previously carried out by a human". Thus, the term *automation* can be applied to any event in which a machine executes a function that is traditionally carried out by a human being. Moray, Inagaki, and Itoh (2000, p. 44) understand automation as "any sensing, detection, information-processing, decision-making, or control that could be performed by humans but is actually performed by machine" (see also Lee & See, 2004). More precisely, Sheridan (2002) argues that automation is best represented by the function it performs (see also Adams, Bruyn, Houde, & Angelopoulus, 2003). Therefore, automation is characterized as a) the mechanization and integration of the sensing of environmental variables (by sensors), b) data processing and decision making (by computers), and c) mechanical action (by motors or actuators) or information action (by communication of processed information to people) (Sheridan, 2002). According to this understanding of the concept, the purpose of automation can be twofold. It can either execute a task that has a direct influence on the environment, or output recommendations based on sensory data about the best decision to aid an operator in processing and integration of environmental information.

The concept of automation can be distinguished from a machine. A complete and permanent reallocation of a function from human to machine is seen as a machine operation (Parasuraman & Riley, 1997). Some functions that formerly have been regarded as automation are part of a larger system and considered a simple machine operation today. Examples from the automotive area for tasks that used to require human involvement are the starter motors or the anti-lock braking system (ABS) for cars. Today, the driver does not have to think about these functions anymore, they are handled by the vehicle for him (Adams et al., 2003).

Machines are, in general, designed to make the life of humans easier, but most of the time, humans are still part of the system as a whole. Those joint systems, where a collaboration takes place between the human and the machine, are called human-machine systems or human-computer systems (Johannsen, 1993). The analysis and optimization of the relationship between the two parties of these systems is the aim of human factors research (Cacciabue, 2004). One objective is the allocation of tasks to define different stages of automation. The classification of levels of automation (LOA) is given in more detail in the following.

Levels of automation

When discussing automated systems, it needs to be specified what level of automation is considered to clarify how much the human operator is still involved in the task. Already in the early Fifties, a first approach to structure the function allocation between humans and machines was proposed by Fitts (1951). It was one of the first publications to suggest that function allocation should take the system's competences into account: some functions can be performed better by the human agent, and conversely, some can be performed better by the machine. This concept lacked the possibility of interaction and shared control, where operator and machine share a task or alternate depending on the situation.

Sheridan and Verplank (1978) introduced another, more flexible approach: a formal taxonomy of automation levels that describes the modes of interaction between human and machine from fully manual operation to fully automated completion of a task (see Table 2.1). Other forms of this scale, with decision process and execution shifting from manual to machine-controlled,

Table 2.1
Levels of automation in man-computer decision-making (adapted from Sheridan & Verplank, 1978).

Automation level	Automation description
1	Human does the whole job up to the point of turning it over to the computer to implement.
2	Computer helps by determining the options.
3	Computer helps determine options and suggests one, which human need not follow.
4	Computer selects action and human may or may not do it.
5	Computer selects action and implements it if human approves.
6	Computer selects action, informs human in plenty of time to stop it.
7	Computer does whole job and necessarily tells human what it did.
8	Computer does whole job and tells human what it did only if human explicitly asks.
9	Computer does whole job and tells human what it did and it, the computer, decides he should be told.
10	Computer does whole job if it decides it should be done, and if so tells human, if it decides he should be told.

can be found in the literature (Endsley & Kaber, 1999; Moray & Inagaki, 1999; Sheridan, 1992; Sheridan & Verplank, 1978; Wei, Macwan, & Wieringa, 1998). The qualitative descriptions of the ten stages illustrate that the human operator is still in charge of the decision making process in levels 1 to 5. In levels 5 and 6, collaboration takes place as the human has to approve the action selected by the computer. Levels 7 to 10 can finally be considered as full automation (Adams et al., 2003; Sheridan & Verplank, 1978), where the human cannot interact with the computer anymore. In this way, the model takes into consideration the processes of decision making and action selection as well as the communication between the two actors.

The taxonomy of Sheridan and Verplank (1978) is limited to the processes of decision making and action selection. A further distinction between different stages of automation is made by Parasuraman, Sheridan, and Wickens (2000). In addition to the levels of automation by Sheridan and Verplank (1978), they describe four stages of information processing that start with processes prior to decision making: a) information acquisition, b) information analysis, c) decision and action selection, and d) action implementation (see Figure 2.1). The stages build up on one another and can be fulfilled either by the human operator or by machine to a certain degree. They are retrieved from several human information processing models that describe (simplified) equivalent stages of sensory processing, perception / working memory, decision making, and re-

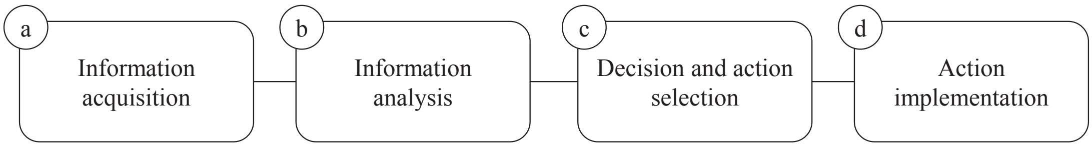


Figure 2.1. Four-stage model of human information processing (adapted from Parasuraman et al., 2000).

sponse selection / implementation (Baddeley, 1996; Broadbent, 1958). According to the model of Parasuraman et al. (2000), the degree to which systems automate these stages of information processing can be used to describe and distinguish them (Popken, 2009).

Classification of automated driving systems

What is valid for automated systems in general also holds true specifically for automated driving systems: it needs to be specified what level of automation is considered. In this context, the key question is translated to how much the human driver is still involved in the driving task. The International Society of Automobile Engineers (SAE) report on levels of driving automation for on-road vehicles provides operational definitions for six different levels of autonomy, ranging from full-time performance by the driver on the one end to full-time performance by an automated driving system on the other end (SAE International, 2014, see Table 2.2). This classification also includes a level of full automation under all road conditions, whereas other taxonomies have been focusing on mode-specific performance of an automated driving system (e.g., definitions by the NHTSA in the United States of America or by the Bundesanstalt für Straßenwesen (BASt) in Germany). While the first levels of automated driving still hold the driver accountable, the SAE taxonomy focuses more on the three higher levels of automated driving, where the automated driving system performs the entire dynamic driving task.

The SAE taxonomy is similar to another one introduced by the German Association of the Automotive Industry (Verband der Automobilindustrie, VDA), except that the six stages are labeled differently. Because the VDA definitions build upon the former BASt classification, high and full automation are adopted and complemented by a sixth stage called “driverless” (Verband der Automobilindustrie e. V., 2015). To avoid confusion, this contribution will always refer to the definitions of the SAE taxonomy of automated driving levels.

Table 2.2
Levels of driving automation by SAE International (2014)

Level	**Name**	**Narrative definition**	**Execution of driving task**	**Monitoring of environment**	**Fallback performance**	**System capability**	BASt / VDA levels	NHTSA levels
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	Driver only	0
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	Assisted	1
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	Partially automated	2
Automated driving 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	Highly automated	3
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	Fully automated	3 / 4
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	(Driverless)	3 / 4

Most of the current systems in cars are still on the level of assistance systems. Examples for assistance systems that aid the driver in controlling the vehicle are the adaptive cruise control system (ACC) with automatic distance control for longitudinal assistance, lane-keeping assistant (LKA) for lateral assistance, and collision avoidance systems like a collision mitigation system (CMS). Decision-aid systems that give recommendations to the driver are for example route navigation systems as well as traffic alerts or traffic sign recognition.

During the last years, more and more European automobile manufacturers and suppliers predicted the implementation of HAD in 2020 and full autonomy from 2025 on (Ziegler, 2013). Meanwhile, automated driving is seen as the self-evident next step in the evolution of driving technology. Figure 2.2 shows the prospect of automated driving development until the year 2025, as suggested by Ziegler (2013).

This work focuses mainly on the levels of conditional and high driving automation, where the automated driving system monitors the driving environment and performs under all roadway and environmental conditions of a specific scenario. All lower modes imply that there are still situations where the human driver is in charge and has to supervise the car to make sure he can respond appropriately to a request to intervene. In conditional and high level of autonomy of the vehicle, the driver does not need to supervise the vehicle anymore. Different from full automation, however, in high automation this is true for a determined context (e.g., on the highway). Based on this assumption, the driver can engage in other tasks during the drive, being completely out of the loop of driving. There might be system limits where the human driver needs to take control within an extended time frame—but once the vehicle is in control, it indicates such system limits to the driver, thus making supervision unnecessary as long as the driver sufficiently trusts the system. The highly automated driving system will at least be capable of stopping the vehicle in a safe state if the driver does not take back control when reaching a system limit. In this context it is often referred to as a fail-operational or fail-safe system, meaning that in the event of a system limit, the system will continue to work in an emergency operation mode, avoid causing any harm and seek a minimal risk state.

2.1.2 Effects of automation

The race is on for automated driving—but how much of the driving task should be automated at all? What are the consequences of using an automated driving system instead of driving manually? Benefits of automation are numerous and are the reason for its widespread implementation. However, drawbacks of automating a task should not be overlooked.

Automation can make our lives easier, more comfortable, and release us from tasks we had to do manually before. Thus, it plays an important role in our daily lives today. Automation is, per

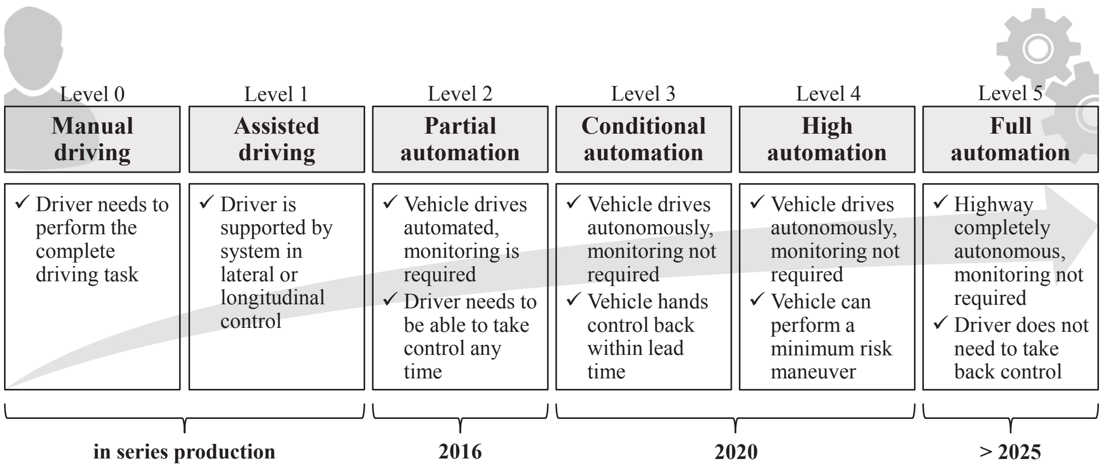


Figure 2.2. The progress of automated driving development from assisted driving to full automation (time frame suggested by Ziegler, 2013).

definition, originally developed to release the human from a task he could do himself. This task either could be done better or faster or simply is less stressful when done by a machine. A human operator who is supported by automation and does not have to do everything manually anymore can pay more attention to other or more important aspects of his task, due to less workload and spare capacity. In addition, human error, that is found to be the main reason for accidents in working areas like aviation or industrial plants (see Billings, 1991; Reason, 1990), can potentially be reduced to a minimum by automating the tasks. It is hoped that automated driving will go along with similar advantages, ranging from increased road safety to more comfort while driving.

However, automation can entail drawbacks. In the industrial sector, the shift to *supervisory control*, first described by Sheridan and Verplank (1978), characterizes the new, different role of the operator inside the human-machine system. He is now no longer an operator, but rather a supervisor of the automated system, and his tasks shift from an active control to passive monitoring (e.g., Richards & Stedmon, 2016; Sarter, Woods, & Billings, 1997; Shen & Neyens, 2014; Walliser, 2011). This has an impact on an issue that is often referred to as *out-of-the-loop performance problem* (Endsley, Bolte, & Jones, 2003; Endsley & Kiris, 1995; Parasuraman, Molloy, & Singh, 1993). It entails three major impairments, which are characterized by Endsley and Kiris (1995) as follows:

- *Vigilance impairment and loss of skills.* Accompanied with the role change to monitoring is a decrease of vigilance. Normally, a loss of vigilance is associated with too-low workload and simple tasks that lead to diminished alertness and can compromise performance. On the other hand, lost vigilance can also occur during complex monitoring tasks, complacent (relying) behavior being the main reason. This can also go along with a loss of skills: as Reason (1990) and Endsley and Kiris (1995) point out, humans are no good supervisors, because they lose their vigilance and their skills over time. While this is not problematic with an accurately working system, in case of a system error this effect can lead to performance impairments (Flight Safety Foundation, 2005). It was shown in several studies that drivers' response to critical events occurs much later when driving automated or with an assistance system compared to manual driving, especially when distracted by a secondary task (Merat & Jamson, 2009; Niederée & Vollrath, 2009; Shen & Neyens, 2014; Young & Stanton, 1997).

- *Loss of situation and mode awareness.* Situation awareness (SA) describes "the perception of the 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. 792). It is thought of as a state of knowledge rather than a process (Ososky, Sanders, Jentsch, Hancock, & Chen, 2014), and consists of the three levels perception, comprehension, and projection, built up on each other. SA is considered an essential construct that drives effective decision-making and performance in dynamic systems (e.g., in the field of aviation) and is an important premise for safe interaction with automation (Endsley, 1993, 1995; Rauch, Gradenegger, & Krüger, 2007). When the active operation and processing of information is degraded to a passive reception of information, this can make a dynamic update of the mental system model difficult. Moreover, an assessment of the situation is impossible when the system assumes the whole task of observing the situation. The driver thus cannot react appropriately to a system failure due to the out-of-the-loop performance problem (Endsley & Kiris, 1995; Kaber & Endsley, 1997).
- *Different feedback mechanisms.* The feedback provided by the system is different when the operator is no longer actually handling the machine. Certain cues might be substituted or even eliminated, especially when it comes to haptic feedback that normally was assessed by physical contact with the machine. Adequate feedback to a human user "is absent far more than it is present" (Norman, 1990, p. 11).

Whenever system boundaries are reached, i.e., when a situation occurs that the system is no longer capable of controlling or which was not foreseen by the system designer, these aspects are important to consider. The human might not be able to do handle this uncommon situation as well: he might be out of the loop of manual machine control or even have lost his skills to operate the machine—and human error can again be made (De Waard, Van der Hulst, Hoedemaeker, & Brookhuis, 1999). Bainbridge (1983) called this phenomenon *ironies of automation*. The human that should be taken out of the equation to minimize errors is again a source of error when needed as a fallback solution.

Benefits and drawbacks of automation use have often raised the questions of what to automate and to what level. Today, the question of what to automate is answered by society itself: thanks to familiarization with more and more advanced technology, nearly everything that can be automated gets automated these days and the question is nowadays merely academic (Endsley & Kiris, 1995). It seems that this trend also applies to automated driving technology, making it necessary to investigate how to automate this part of our lives best.

2.1.3 Challenges for automated driving systems

As has been explained, making automated vehicles become a reality on the street is not only a question of technical feasibility any longer. The novelty of self-driving vehicles adheres to several challenges that need to be mastered in order to advance this topic. In the beginning of this development, the technical feasibility had to be addressed and a lot of effort was put in research and development of concepts to drive automated, pursuing different strategies regarding the application of technical equipment. Now that the main technical issues are solved and the first vehicles are driving automated already, other topics are getting more pressing. To provide a more complete picture of factors that can influence the use of automated driving systems next to trust, arising challenges are briefly discussed here. One important question that needs to be solved before automated vehicles can be allowed for public use and be sold to customers concerns legal liability issues. How can responsibilities be clarified if, for instance, an autonomously driven car gets involved in an accident?

Legal issues

It is crucial to pave the way for automated driving by overcoming legal barriers and regulatory hurdles that are still in the way of this technological revolution. First steps have already been taken, initiating a change of the Vienna Convention on Road Traffic of 1968 (Articles 8 and 13, United Nations Economic and Social Council, 1968). The convention forced drivers to have control over his moving vehicle at all times, leaving no room for systems that could replace the driver in this matter. To provide a legal basis that allows automated driving to be implemented, it was agreed to introduce an amendment on Article 8 of the Vienna Convention. Since March 2014, systems are allowed to control the vehicle as long as a driver is present and able to override it or switch it off at any time (United Nations Economic and Social Council, 2014).

The level of driving automation is highly relevant for legal and liability issues (Gasser, 2012). Especially for higher automation levels, Gasser (2012) recommends an analysis on a national level. First attempts are made to regulate the use of automated vehicles, one example being the report of the Ethics Commission by the German Federal Ministry of Transport and Digital Infrastructure (2017). The report comprises twenty propositions on automated and connected driving. It claims rules for automated vehicles that shall govern how the automated systems perform in critical collisions. One fundamental tenet is that material damage should be preferred compared to personal injury; another principle says that no distinction between humans should be made, e.g., based on height or age.

Mixed traffic

Of course it will take a long time until vehicles will be on the road that are not controlled by the driver anymore. And it will take even longer until most of the vehicles are equipped with such a system. Until then, the development of automated vehicles needs to consider a mixed-traffic environment. Vehicles equipped with new technologies need to show consideration for both other automated vehicles as well as conventional vehicles driven by humans. This communication can be difficult for an automated vehicle, because it cannot rely on eye contact, gestures, and others communication channels in case of a misunderstanding or communication difficulties. On the one hand, it needs to understand what other road user want to signal them, and on the other hand, it needs to convey its own intent to the others. Thus, not only the interior HMI is of importance—the outward communication needs to be designed with just as much care (see for example project interACT (2017), funded by the European Union).

An automated vehicle will necessarily be programmed to drive defensively and smoothly to avoid disturbance for its passengers. It needs to be programmed to comply with all traffic rules and speed limits. This could lead to problems, the automated vehicle becoming a hindrance to conventional traffic that flows faster despite a speed limit. Situations can arise that demand a different behavior than usual, and that an automated vehicle cannot handle as creatively as a human driver. This could also potentially lead to people challenging the system to see how the system reacts. These challenges need to be faced as long as there are both automated vehicles and traditional ones on the road.

Standardization

To support the exchange of information between the automated vehicle and other cars, efforts are being made towards the implementation of communication standards for vehicle-to-vehicle and vehicle-to-infrastructure communication. In Germany, for example, the Federal Ministry of Transport and Digital Infrastructure established the round table "Automated driving" to discuss requirements and framework conditions together with government agencies, federal states, the insurance industry, research institutes, and automotive manufacturers. With the project “Digitales Testfeld Autobahn”, the ministry is furthermore providing a testing environment for new communication systems and technologies to evaluate their potential long-term benefits (German Federal Ministry of Transport and Digital Infrastructure, 2015). Other states in Europe initiate similar projects, and projects like EPoSS (European Technology Platform on Smart Systems Integration) aim at harmonizing the different initiatives in Europe to agree on common standards (European Technology Platform on Smart Systems Integration, 2015).

Ethical dilemmas

When all legislative issues are solved and a comprehensive framework is developed, what still remains is the ethical question of how an automated driving system should behave. Dilemma situations are posed, leaving a future automated driving system a terrible choice between two fatal accidents (Maurer et al., 2015). However unlikely these dilemmas are, they illustrate how difficult the programming of such a vehicle is—an ethically correct answer might not even exist. However, thinking about these dilemma situations can help to develop strategies that are taking ethical questions into account for further development.

Agreements need to be reached concerning responsibilities and insurance payments in order to be ready for all eventualities. Also, privacy issues are arising when the vehicle is able to send and receive information on its own to communicate with other road users or a network. Finally, it might be necessary to discuss the overall ethical question of what the car needs to be able to handle and how it should react to certain situations. Companies aiming to address the topic of automated driving will have to proactively engage with government legislative groups in order to clear the way for this technology.

2.2 Trust in automated systems

In this section trust is introduced as an essential construct for the use of HAD systems. Trust was found to be one of the major conditions for reliance and use of automation (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). Although trust is mainly considered a psychological state that is relevant for interaction between people, it is more and more adopted in the context of automation as well. When approaching the psychological concept of trust in such a different context, a precise definition and applicable models need to be utilized.

2.2.1 Definition of trust

Trust is generally considered a mental state, similar to an expectation about a certain competent behavior of another party (De Vries, 2004). Trust is seen as a multi-dimensional and dynamically changing concept (Atoyan, Duquet, & Robert, 2006; Dzindolet et al., 2003), consisting of various different components. Regardless of the field of research, in both automation and interpersonal trust literature it is stated that the basis of trust consists of cognitive as well as affective characteristics (Adams et al., 2003).

When defining trust in systems or in automation, it is often referred to trust in other humans, especially to interpersonal trust as a more specific, interaction-related kind of trust (De Vries, 2004). The question is how well concepts of trust in relationships translate to trust between a human and a machine. Madhavan and Wiegmann (2007) provide a comparison of trust in a human adviser and trust in a decision aid system. Differences in the assessment of the counterpart and in monitoring strategies are described. For example, it is proposed that the human operator has a certain response bias, depending on the supposed features of the interaction partner. He will, for instance, expect a human to behave flexible and adapt to a situation, whereas a machine is supposed to react in an invariant way. The assessment of behavior of the interaction partner is filtered through the human's cognitive schema, i.e., an assumption of perfection for the automated system and an expectation of imperfection for the human counterpart (Dzindolet, Pierce, Beck, & Dawe, 2002). The monitoring strategy is adapted to this assumption, also influenced by the self-confidence of the operator. Combined with the primary basis of trust judgments, this leads to a resulting assessment of trust (or distrust) in the interaction partner (Madhavan & Wiegmann, 2007). Interpersonal concepts have been proven to be related to trust in machines (e.g., Muir, 1994), but seem to not be completely interchangeable, as the comparison of Madhavan and Wiegmann (2007) shows (Lewandowsky, Mundy, & Tan, 2000). In fact, Adams et al. (2003) argue that there are profound differences between the two concepts, i.e., trust in automation being unidirectional and trust in a human being driven by the aim of earning the other's trust. In this work, the term *trust* refers to trust in automated systems or human-machine trust if not specified otherwise. Nevertheless, some of the research papers presented here are originally conducted in a different context, but still can be compared or used to further describe the concept of trust in automation.

So what does trust mean in the context of HAD and how can it be defined? When interacting with automated systems in dynamic, time-critical situations, humans are likely to have difficulties with perceiving and processing all necessary information to manage the situation properly (see Moray et al., 2000). In that case, they have to act under uncertainty, not knowing all factors relevant for interpreting the situation (Rajaonah, Anceaux, & Vienne, 2006). Especially when automation assumes most of the task the human normally does, the human is left without further information about the handling of the situation. Rajaonah et al. (2006, p. 101) conclude that "human-machine systems require an internal mechanism that will allow operators to reduce the feeling of uncertainty and risk related to the possible consequences of their decisions, this mechanism being trust." Depending on the field of research and the focus of investigation, there are various definitions of trust that can be found in literature. An often cited and generally accepted organizational definition is brought up by Mayer, Davis, and Schoorman (1995, p. 712). In their definition, they describe 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". Applied to automated driving, trust represents a relevant construct because the driver's role shifts from an active agent who directly controls the vehicle's action to a passively monitoring one. Even though the driver, in principle, can always intervene and take manual control, the basic concept of HAD does not require such intervention during routine operation. In case of a detected system limit, the driver is informed by a salient signal prompting him to take over manual control of the vehicle within a defined time frame. Trust in a decision aid is defined by Madsen and Gregor (2000) as the extent to which a user is confident in, and willing to act on the basis of, the recommendations, actions, and decisions of an artificially intelligent agent. In a similar manner, Lee and See (2004, p. 54) define trust as "the attitude that an agent will help achieve an individual's objectives in a situation characterized by uncertainty and vulnerability". This is especially true for today's highly complex systems, where uncertainty can arise regarding whether the system will work well in a certain situation. This definition is widely adopted in the context of automation, and is thus also the foundation this work builds upon.

During highly and fully automated driving, drivers are not in charge anymore and do not need to supervise aspects of the driving task. However, whether drivers truly rely on the automation depends on their trust in the efficacy of the automation (and, in comparison, on their belief in their own ability to control the vehicle) (Lee & Moray, 1994). These constructs play a major role in dynamic *allocation of function*, when drivers can decide to use or not to use an automated system (Lee & Moray, 1992; Moray et al., 2000).

2.2.2 Effects of trust

Several other psychological concepts are related to the topic of trust in automation and can sometimes be confused with it. Their relation to trust is described in the following.

Reliance and compliance

Trust is, as the definition says, an ever-changing attitude, whereas automation dependence (like reliance and compliance) is understood mostly as the (potentially) resulting behavior (Wickens, Hollands, Banbury, & Parasuraman, 2016), reflected for example in use or disuse of a system (Lee & Moray, 1992). The two constructs are defined as follows: “Compliance is what the operator typically does when the automation diagnoses a signal in the world, whereas reliance is what the operator does when the automation diagnoses noise in the world” (Dixon, Wickens, & McCarley, 2007, p. 564). A conceptual framework developed by Popken (2009) describes human adaptation to automation using the concepts of reliance and situation awareness (Figure 2.3). In this model, reliance is seen as an observable outcome of trust (Wickens et al., 2016), and trust can be understood as an attitude preceding reliance (in line with the model of trust by Lee and See (2004), see Section 2.2.3). Complacency is treated as another attitude toward an automated system affecting reliance. Other factors in humans’ adaptation process that influence the intention to rely according to Popken (2009) are energetic processes like vigilance, arousal, and mental workload, as well as motivational processes like mental effort regulation.

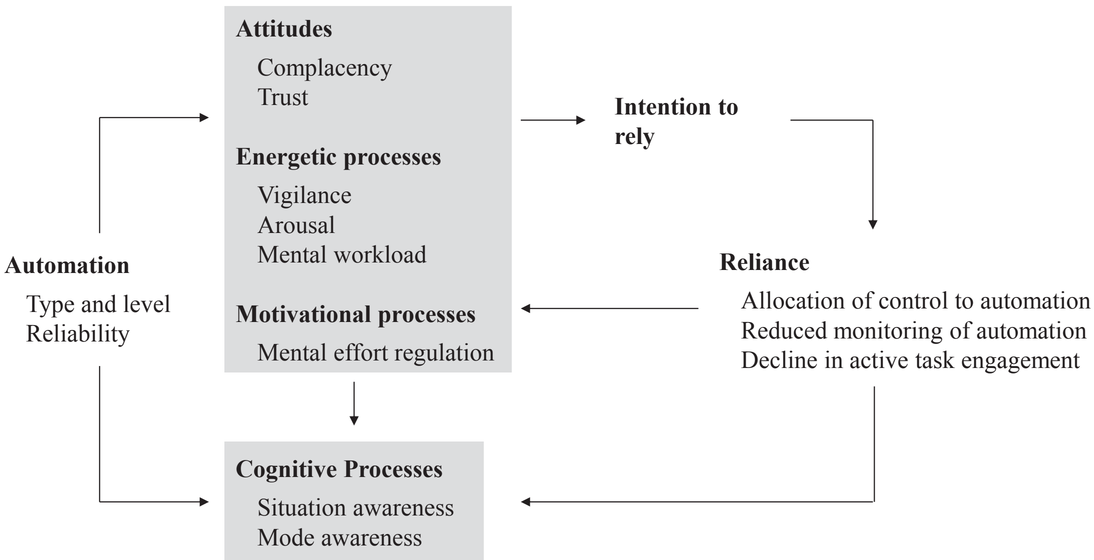


Figure 2.3. Model of reliance (adapted from Popken, 2009).

A diversity of experiments have revealed a relation between trust and reliance. Sheridan and Hennessy (1984), Zuboff (1988), as well as Lee and Moray (1992) all found a positive correlation between trust in a system and its use. A link between trust and reliance was also reflected in results of Muir (1994) and Muir and Moray (1996). Trust was found to be one of the major determinants and the strongest predictor for system use (Lee & See, 2004; Masalonis & Parasuraman, 1999). An operator might use a system that is not reliable, simply because he trusts it. Conversely, even if an automated system operates reliably, a human operator might not rely on it if he beliefs the system is not trustworthy (Parasuraman & Riley, 1997). Muir (1994, p. 1911) thus asserts: "The expectation of technical competence is probably closest to our intuitive understanding of what it means to trust a machine." Regarding automation use or reliance on automation, Muir (1987, p. 1906) states: "When human supervisors allow automation to control a process, we may infer that they trust that automation, to some extent at least". Thus, reliance is understood as the concrete observable behavior of allocating control to an automated system, resulting from the psychological state of trust (Beggiato & Krems, 2013).

However, Atoyan et al. (2006) have correctly reminded that reliance does not necessarily indicate a high level of trust—a person can have other reasons (for example a high level of workload) for relying on an automation that is not trusted. This observation was also confirmed by Riley (1994). He was able to demonstrate in four different experiments the influence of automation reliability, task uncertainty, and risk on the decision to rely on an automated system. It becomes clear that the relation between trust and actual behavior is still discussed controversially and cannot be safely assumed. On this issue, Chancey, Bliss, Liechty, and Proaps (2015) have emphasized that a willingness to be vulnerable in the sense of trusting a system does not require any risk-taking or an actual response behavior. The authors thus argue that a perceived risk needs to be involved in the situation to make trust a strong source of the response behavior.

System use

A successful interaction between a human operator and an automated system requires an adequate allocation of control. Inappropriate allocation of control can result in one of the following categories described by Parasuraman and Riley (1997).

Misuse. Misuse of an automated system can occur when operators rely on automation in situations it should not be trusted because it is not reliable. Misuse is defined as "overreliance on automation" by Parasuraman and Riley (1997, p. 230). This overtrust in the automation can lead to insufficient supervision of the outcome of an automated task, resulting in unnoticed malfunctions or errors. Researchers differentiate between two errors occurring due to inappropriate allocation of control in a decision aid system (Alberdi, Strigini, Povyakalo, & Ayton, 2009; Parasuraman & Manzey, 2010; Popken, 2009; Skitka, Mosier, & Burdick, 1999):

- *Errors of omission* can take place when the automation fails to report an error in the system, and the operator does not monitor the system in an adequate way to notice the miss. Errors of omission thus arise from undue reliance.

- *Errors of commission* occur when the automation hands out a wrong advice that is incorrectly followed by the operator without further examination. These errors can happen due to an operator's inadequate compliance.

In the context of HAD, no automation failures are expected. In this level of automation, redundancies are implemented to enable the system to function until it can hand control back to the driver.

Abuse. If the automation is used in situations it was not originally designed for, an inappropriate application of automation takes place, which is referred to as abuse (Parasuraman & Riley, 1997). It can lead to system failures and a reduced performance of an automated system. This could be the case if an automated driving system is activated in a situation it was not designed for (e.g., a highway automated driving system that is activated in the city). Ideally, an abuse of this kind is prevented by the system itself.

Disuse. In contrast to misuse, disuse describes an underreliance on or underutilization of automation, although the reliability of the automation is high (Parasuraman & Riley, 1997). In that case, warnings or advises of the automated system might be ignored. It might as well be a consequence of false alarms that diminish trust in the system. Despite the high reliability of nowadays systems, false alarms still do occur, causing the human operator to rely less on the automation. Future alarms might even be ignored from that moment on, which is called *cry-wolf effect* (Breznitz, 1984).

Experiencing an automation can lead to a reaction of two kinds. Muir (1994) describes *mistrust*, a false trust although the automation's performance is poor, in contrast to *false distrust*, where the automation is not trusted although it's performance is high (see Table 2.3).

Table 2.3
Interactions of operator's trust in and use of automation with the quality of the automation (adapted from Muir, 1994)

Operator's trust and allocation of function	**Quality of the automation**	
	'Good'	**'Poor'**
Trusts and uses the automation	*Appropriate trust*, optimize system performance	*False trust*, risk automated disaster
Distrusts and rejects the automation	*False distrust*, lose benefits of automation, increase operator's workload, risk human error	*Appropriate distrust*, optimize system performance

Both over- and underreliance can be critical when it comes to automation use, as Figure 2.4 shows. Overtrust, on the one hand, gets the operator to rely on the automation more than is appropriate. This out-of-the-loop problem can in turn cause the operator to have low confidence in his own skills, thus using the automation more. With undertrust, on the other hand, the operator is likely to not use the automation and can therefore not gain experience with the system (Muir, 1994). Naturally, an automated system can only prove itself worthwhile when activated, and the human can only gain confidence in the system when he uses it.

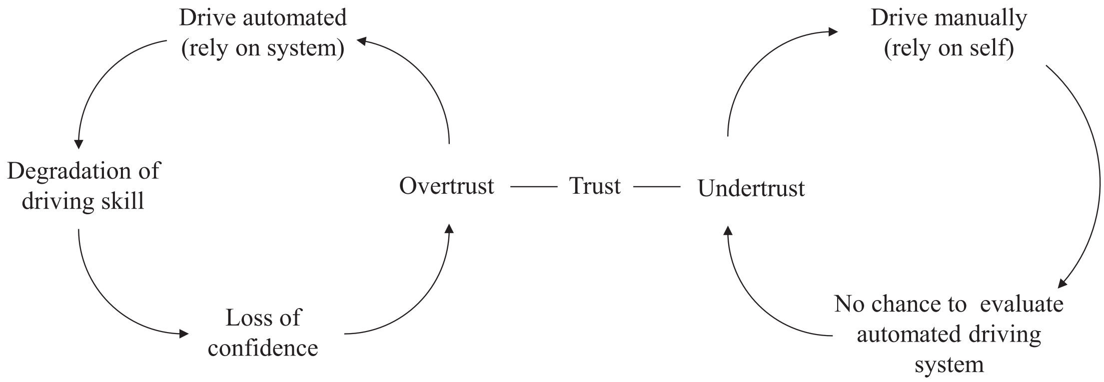


Figure 2.4. Cycle of trust in the context of automated driving (adapted from Llinas, Bisantz, Drury, Seong, & Jian, 1998).

Automation bias and complacency

In both types of misuse errors (errors of omission and errors of commission), according to Mosier and Skitka (1996) as well as Mosier, Skitka, Heers, and Burdick (1998), humans tend to use cues or aids from an automation as a heuristic replacement for their own information seeking, perception, and processing. This effect in the use of automated aids and decision support systems is referred to by researchers as *automation bias*. Likewise, research of Dzindolet et al. (2003) proves that humans expect other human interaction partners to be outperformed by automated decision aid systems.

The issue of not monitoring an automated system or verifying decisions of a system by checking the raw information sources is also called *complacency* (Parasuraman & Riley, 1997; Parasuraman, Sheridan, & Wickens, 2008; Singh, Molloy, & Parasuraman, 1993). A National Aeronautics and Space Administration (NASA) report defined complacency as “self-satisfaction which may result in non-vigilance behavior, based on an unjustified assumption of satisfactory system state” (Billings & Cheaney, 1981, p. 31). Wiener (1981, p. 117) suggested a similar definition, describing complacency as “a psychological state characterized by a low index of suspicion". A positive bias towards automated systems leads to an overly high expectation of automation performance and is assumed to be the reason for this phenomenon (Dzindolet et al., 2002) (also called *perfect automation scheme* by Bahner (2008)). Findings of Dzindolet et al. (2003) as well as Madhavan and Wiegmann (2007) also support the existence of such a positive bias towards automation.

The research of Parasuraman and Manzey (2010) gives more insights in the development of complacency as a construct similar to trust. The integrated model of complacency and automation bias can be seen in Figure 2.5. Both complacency and automation bias are assumed to result from dynamic interactions between personal, situational, and automation-related characteristics. Former research viewed complacency and automation bias as entirely independent constructs resulting from an automation design. However, Parasuraman and Manzey (2010) point out that the two constructs represent different manifestations of an automation-induced phenomenon. *Complacency* is seen as an attention allocation strategy, while *automation bias* is understood as the outcome of this strategy, namely errors of omission and commission.

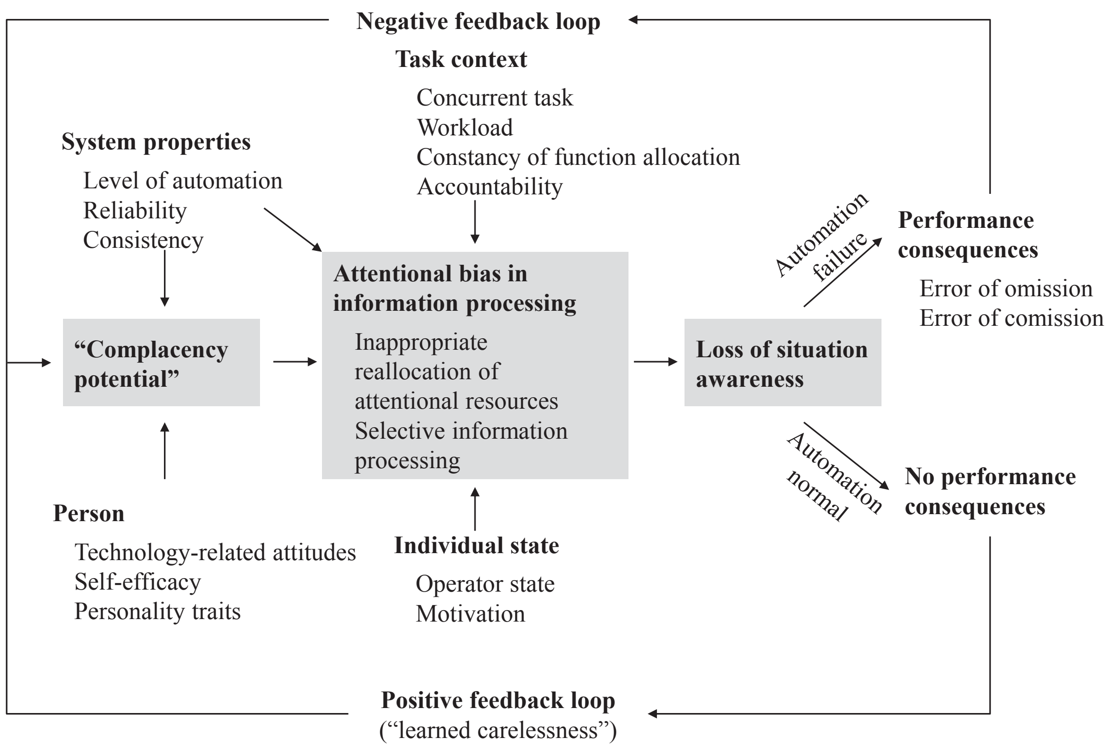


Figure 2.5. Integrated model of complacency and automation bias (adapted from Parasuraman & Manzey, 2010).

While trust in automation results in effects that need to be accounted for in human-machine interaction, the factor itself is influenced by a diversity of variables. The next Section 2.2.3 describes which factors can influence trust in an automated system.

2.2.3 Dimensions of trust

A variety of different internal characteristics and experiences of the human as well as characteristics of the system, the situation, and the environment have shown to play an important role for developing trust in an automated system. Overviews of relevant factors can be found in Lee and See (2004), Merritt and Ilgen (2008), or Hancock et al. (2011). A recent work of Hoff and Bashir (2015) has synthesized the current state of the art to a three-layered model of trust, including the three layers dispositional, situational, and learned trust.

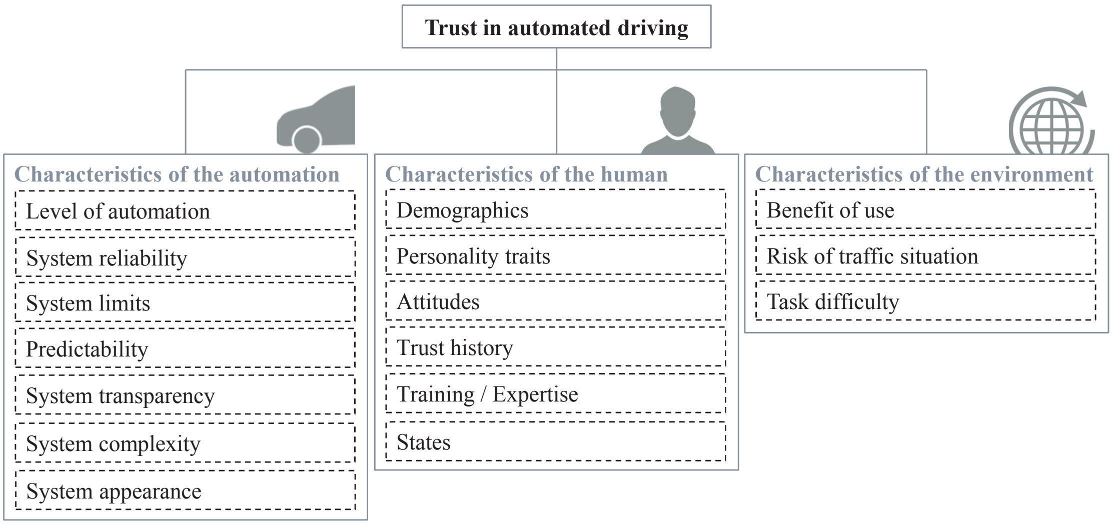


Figure 2.6. Overview of factors influencing trust as proposed by literature (see Hancock et al., 2011; Merritt & Ilgen, 2008).

The factors are discussed here in order to provide an overview of the whole scope of the topic (adapted from Hancock et al. (2011), see Figure 2.6). In general, those constructs are often divided into two (or sometimes three) groups, the first one looking at the properties of the machine, the other one at the characteristics of the human. Sometimes a third group is considered to describe the interaction between the first two groups or situational and environmental factors. Merritt and Ilgen (2008), for example, describe that trust can be affected by characteristics of the automation, the operator, and the context.

Similarly, the objective of research on trust in automation conducted by Cohen, Parasuraman, and Freeman (1998) was to develop a framework for understanding trust in decision aids. The argument-based probabilistic trust (APT) model they generated illustrates the variation of trust depending on the user's personality, the characteristics of the automated system, the specific situation, and expertise level of the user (see Masalonis & Parasuraman, 1999). In the following, the groups of factors are presented in more detail and accordant research results are aggregated.

Characteristics of the automation

Some key issues related to trust in an automated system are based on the characteristics constituting the machine. The most relevant ones are described briefly in the following section.

Level of automation. The different levels of automation, ranging from merely assisted to full autonomy, can have an impact on trust in the system. For example, the results of Walliser (2011) suggest an influence of the level of automation on an operator's ability to calibrate trust. Also, there is an effect on performance in case of an error. With a higher level of automation, it was found that participants had significantly longer response times to a system failure than with a lower level of automation (e.g., Niederée & Vollrath, 2009; Shen & Neyens, 2014, see Section 2.1.2). While system errors need to be taken into account when interacting with lower levels of automation, they are not to be expected in higher levels (high and full automation). Still, undesired or unexpected reactions of an automated system may occur, leading to a feeling of failure. Thus, research results on system reliability and system limits may be important for the consideration of trust in highly automated driving.

System reliability. Trust in automation is, for the most part, dependent of the performance of the automation. When automation is reliable, trust in the system is higher and the automation is more likely to be used (Muir, 1994; Muir & Moray, 1996). System reliability, sometimes also referred to as competence (Muir, 1989), is the consistent good performance of the system. Associated with this performance of the system is the resulting user trust, with users relying only on automation that is trusted more than their own abilities to operate a system (Merritt, Heimbaugh, LaChapell, & Lee, 2013, see also Section 2.2.3). Reliability is able to shape trust in a system depending on the user's expectancy and the actual reliability (De Vries, 2004; Kazi, Stanton, Walker, & Young, 2007; Lee & Moray, 1992, 1994; Moray et al., 2000). Reliability seems to determine reliance as well: Buld et al. (2002) found longer reaction times to automation failures when reliability was high. Research results reported by Lee and Moray (1992) describe that there is also some evidence that only the most recent interaction with automation impacts trust modulation (Adams et al., 2003). In addition, the stability of the system's reliability is quite important for its predictability and trustworthiness: a stable performance with a non-fluctuating reliability makes the automation more predictable and thus more trustworthy (Muir & Moray, 1996).

System limits. System limits or failures of an automated system as discrete manifestations of low reliability of a system have been reviewed a lot, providing insight in their influence on trust. System faults in general undermine trust in the system. This is dependent on the magnitude of the failure as well as on their variability, as results by Lee and Moray (1992) show. Muir and Moray (1996) found out in particular that several small errors had a more severe impact on trust than one larger error. Also, *false alarms* were found to mostly affect operator compliance, whereas *misses* seem to affect operator reliance (Dixon et al., 2007). Moreover, it has been shown that trust suffers a great drop and recovers only slowly from errors, even if performance is reestablished rather quickly after the error (Lee & Moray, 1992; Ma, 2005). As mentioned before, an important finding of Riley (1996), Dzindolet et al. (2003), as well as Beggiato and Krems (2013) revealed that knowledge about faults in advance can diminish their effect and thus be more important than the actual performance of the system (see also Lee & Moray, 1992, 1994; Muir & Moray, 1996). If errors are predictable, a system might be used and trusted despite the errors. Even continuing small errors can be compensated for if the operator understands the system's behavior and boundaries (Lee & Moray, 1992; Ma, 2005; Muir & Moray, 1996). Contrarily, a discrepancy between operator's expectations and the system performance can have a negative effect on trust—even when the automation functions as provided (Lee & See, 2004). Another interesting result was revealed by Madhavan, Wiegmann, and Lacson (2003), indicating that task difficulty can guide the level of trust in an automated system as well. Single failures of an automated system are especially harmful for trust when the automated task is perceived as easy. Furthermore, trust in the system is prone to the primacy-recency-effect: when automation reliability is low in the beginning of the interaction, the system might not be trusted enough to use it further on (Atoyan et al., 2006). A future HAD system may inherently have a high reliability and may be designed in a way that sensor redundancies absorb potential errors—yet, there might be gradations, e.g., in terms of the number of takeover requests or the stability of driving performance. These gradations could be perceived as a low reliability even if the system can handle all situations within the system limits.

Predictability and system transparency. Predictability of a system is related to its reliability and the consistency of the system's performance. Thus, the perceived dependability of the automation can impact trust as well. The expectation of predictability is thought of by Rempel, Holmes, and Zanna (1985) as well as Muir (1987) as one of the major factors influencing trust (see Section 2.2.4 for a more detailed description). Predictability also goes along with the system's transparency (Ghazizadeh, Lee, & Boyle, 2012): corresponding with the idea of observability, the behavior of the automation can only be predicted when the system's actions are comprehensible and rationally explainable to the user. Ososky et al. (2014, p. 1) have defined system transparency as "the degree to which a system's action (. . .) is apparent to human operators and/or observers". This might be attained by designing an automation that acts in a manner similar to how the human might act, or by creating a system that can communicate information to the human in order to explain its intents and actions (Ghazizadeh et al., 2012; Sarter et al., 1997; Seppelt & Lee, 2007). Norman (1989) generally attributes automation-related accidents to the erroneous assumption that information does not need to be provided on tasks the system is doing autonomously. This can lead to fatal accidents in the event of system boundaries, as the operator then cannot identify the situation and react appropriately. "Providing adequate feedback under automation to keep the operator informed, yet not overloaded, may be a formidable challenge for designers of future systems" (Endsley & Kiris, 1995, p. 384). Based on this assumption, system feedback and thus system transparency is just as necessary when a task is automated, maybe even more important than during manual execution. A system that is transparent is explaining the processes underlying the automation, thus facilitating the understanding of the functioning or malfunctioning of the automation. Simpson and Brander (1995, p. 77) describe that a system can only be trusted if it demonstrates technically competent role performance and helps the human to predict the pattern of its accuracy. Research has established that transparency enables mental models to be created or updated based on information about the system, thus avoiding *automation surprises* (Sarter et al., 1997) and helping to explain system boundaries or failures. Consequently, self-explanation abilities of the automation can help for example during system errors. Experiments of Dzindolet et al. (2003) revealed that participants were relying more on a decision aid system and trusted it more when a reason was provided regarding why the aid might err (thus increasing responsibility of the system). This might be due to a distinct mental model that can evolve because of these explanations (Adams et al., 2003, see also Section 2.2.3).

System complexity. When the *complexity of a system* is low, peoples' reliance on the automation is only loosely coupled to their trust in the system. This relation is getting more important with higher complexity of an automated system, and people who trust the system are more likely to use it (Lee & See, 2004).

System appearance and reputation of system designer. Regarding the characteristics of the automation, an overview of relevant factors is provided by Söllner, Hoffmann, Hoffmann, Wacker, and Leimeister (2012), looking at formative first- and second-order factors for trust development in information technology artifacts. The first-order factors *performance* (what is the system doing?), *process* (how is the system working?), and *purpose* (why was the system developed) are derived from the trichotomy of trust described by Lee and Moray (1992) (see Section 2.2.4), and used to structure relevant second-order characteristics of an automation. When attempting to use the same approach with factors relevant in the context of automated driving (see Figure 2.6), the first-order factor *performance* could be used to subsume the level of automation, system reliability, and failures of the automation (similar to Söllner et al., 2012). According to this pattern, the factors transparency, predictability, complexity, and appearance of an automated driving system are more related to the first-order factor *process*. The factor *purpose* is more difficult to define for the context of automated driving systems, as it is related to the question why the automation was developed. It is assumed that for this context, most relevant second-order factors are related to situational conditions.

Characteristics of the human

For trust in an automated system to develop, also a variety of internal characteristics and experiences of the human play an important role. Personal traits as well as current states can be relevant characteristics. Characteristics of the operator that can have an influence on specific layers trust are described in the following paragraph (see model of Hoff & Bashir, 2015, in Section 2.2.4).

Demographics. Dispositional trust, as a relatively stable construct, can nonetheless be subject to changes. Regarding demographic characteristics, four primary sources of variability were revealed to be important in this most basic layer of trust—culture, age, gender, and personality (Hoff & Bashir, 2015). Studies found cultural differences in how people interact with automation (Heimgärtner, 2007) and how much they trust it based on their cultural background (Hoff & Bashir, 2015). Other research was able to show that people of different ages rely on automation differently. For example, results of Sanchez, Rogers, Fisk, and Rovira (2011) indicate that older users rely less on an automated system in the beginning of the interaction and their level of trust is more appropriate to reliability changes of the system.

Personality traits. Merritt and Ilgen (2008) found evidence that trust in machines is linked to aspects of personality like *extraversion*. This corresponds to similar relationships found between extraversion and interpersonal trust, and has theoretically been explained by the definition of extraversion. Extraverts have a general tendency to be more sociable and open to others than introverts, which also seems to transfer to technical systems they interact with (McBride & Morgan, 2010; Merritt & Ilgen, 2008). For example, extraversion was found to be positively related to initial trust in an automated system. Experiments by Merritt and Ilgen (2008) showed that extroverts' initial trust in a system is higher than the initial trust of introverts, and is affected more when the performance of the automation is worse than expected. As trust is relevant in situations of uncertainty and loss of control, a person's *desire to be in control* (Burger & Cooper, 1979) is expected to influence general trust in automation (Gebhardt & Brosschot, 2002). Burger and Cooper (1979, p. 1) define it as the "motivation of being able to control the events in one's life" and describe it as a strong human need. Studies on human-robot interaction provide findings on the relationship with acceptance, indicating that humans prefer a robot that asks permission before acting (Okita, Ng-Thow-Hing, & Sarvadevabhatla, 2012). Another personality factor associated with trust in automation is the *tendency to take risks* or risk propensity. Sitkin and Pablo (1992, p. 12) define it as "the tendency of a decision maker either to take or to avoid risks". Muir (1994) suspected that the tendency to take risks may affect the development of faith as one stage of trust in automation. This assumption was later supported by Desai et al. (2012), who reported that participants who were willing to take risks also tended to trust a robot more than less risk taking individuals during a robot control experiment. A concept related to trust, but distinct from it, is the construct of self-confidence or *self-efficacy*. Bandura (1997, p. 382) explains that "perceived self-efficacy refers to belief in one's agentive capabilities, that one can produce given levels of attainment. A self-efficacy assessment, therefore, includes both an affirmation of a capability level and the strength of that belief." In contrast to self-efficacy, the author regards confidence as a nondescript catchword not embedded in a theoretical system. Despite the judgment of Bandura (1997), relevant results regarding self-confidence are reported in the following, as they are directly related to the construct of self-efficacy. Muir (1994, p. 1915) argues that "an individual who makes a prediction may associate a particular level of certainty, or confidence, with the prediction. Thus, confidence is a qualifier which is associated with a particular prediction; it is not synonymous with trust." Even more, Numan (1998) considers *confidence* as an expectation based on evidence without any uncertainty, *trust* based on partial evidence and *faith* based on no evidence whatsoever (De Vries, 2004). Research of Lee and Moray (1994) found that people with high perceived self-confidence are more likely to develop high trust in automation. Results of numerous studies showed the interdependence between self-confidence and automation use (Lee & Moray, 1992, 1994; Lewandowsky et al., 2000). These findings indicate that automation is used instead of manual control if the trust in the system exceeds the operator's self-confidence, and the other way around. It is the difference between trust and self-confidence that is the actual predictor of automation use (Masalonis & Parasuraman, 1999). Self-confidence, in contrast to trust, is not affected by system reliability. Luhmann (2000) argues that trust presupposes a situation of risk, and confidence does not. This characterization has even been extended, with trust depicted as having more of an affective component and confidence as being rather cognitive (Madsen & Gregor, 2000).

Attitudes. Technology affinity is defined by Karrer, Glaser, Clemens, and Bruder (2009) as a positive attitude and enthusiasm over technology that has a positive effect on knowledge and experience with technology. Besides the personality traits, *acceptance of technology* has been suggested to be linked to the strength of general trust in technology (Ghazizadeh et al., 2012). Such a general attitude of a person towards technology can have an influence on how initial trust in an automated system is pronounced and on how it develops while using the system.

Trust history and experience. Experience with automation is likely to affect trust, as expectations regarding the automation are shaped based on the experienced reliability of the automation. Not only were Sanchez et al. (2011) able to show that depending on the level of experience with the system, the impact of low system reliability varies. Manzey, Reichenbach, and Onnasch (2012) furthermore discovered that an overall system experience with a negative connotation entails much stronger effects compared to a positive feedback loop. Thus, the amount and kind of experience with a system is an important factor influencing reliance in it.

States. Rather than a trait, *stress*, *mental workload*, and *confidence* are important states of a person, potentially affecting reliance on automation. The more different tasks a human has to fulfill that can be automated, the higher will be automation use (Parasuraman & Manzey, 2010). Furthermore, Merritt (2011) describe the affective influence of emotions (e.g., happiness) on trust and liking for automated systems.

Characteristics of the situation or the environment

Situational factors can play an important role, not necessarily influencing trust directly, but determining the extent to which trust influences behavior towards automation (Hoff & Bashir, 2015). For example, Hoff and Bashir (2015) describe how environmental conditions like the novelty of a situation, the operator's degree of decisional freedom, or the operator's ability to compare automated to manual performance can promote stronger relationships between trust and reliance. When the situation allows for the human to check on the automation and enables him to verify the correctness of the system's behavior, trust will have a greater influence on reliance on the system. Additionally, the perceived benefits and risks of using an automated system, as well as task demands and the current workload of the human are influential.

Risk and benefit. The *benefit* of using an automated system naturally has an influence on automation use. If the advantages of using an automated system are marginal, a human will not feel compelled to rely on the system. If, on the other hand, he feels that using the automation would have immense advantages for him, he will be more likely to use the automation even if he lacks confidence in the automated system. A factor that can have a direct effect on trust is the *riskiness* of the situation. As the construct of trust is becoming especially relevant in situations of uncertainty, the level of risk immanent in a certain situation has an impact on the resulting trust. An increased level of risk on hazards leads to decreased trust and use of automation (Perkins, Miller, Hashemi, & Burns, 2010). Reliance on an automated system in a situation of high risk can indicate a larger amount of trust (Muir, 1994).

Task difficulty and situation complexity. Internal processes can depend, in parts, on the situation or the task at hand. As the objective of automated systems is to release the human from doing strenuous or parallel tasks, reliance on automation is dependent on the current task demand and the complexity of the given situation. Whenever task demand is higher than can be carried out by the human operator, he will rely more on the automation than when he has the capacity to monitor and cross-check the automation (Parasuraman & Manzey, 2010). Also, whether people rely on an automated system depends on their perception of the efficacy of the automation and their perceived own ability to master the task at hand (Lee & Moray, 1992, 1994; Moray et al., 2000).

To summarize the insights gained through literature research on determinants influencing trust in an automated driving system, it can be noted that several groups of factors seem to be relevant. They are not limited to characteristics of the automation, nor do they solely stem from the human character. Both areas are key factors that need to be considered when exploring trust in an automated system, and they are furthermore influenced by the situation and the environment the interaction is taking place in. To structure those various constructs in an appropriate way, lots of research was conducted on models of trust in automation. Starting in 1985, numerous models were developed to make the construct of trust more understandable and be of use for further research on this topic. The next Section 2.2.4 will give a short overview of existing models of trust, focusing on models suitable for the context of automated driving.

2.2.4 Models of trust

A lot of research already looked into the topic of trust in automation. Some of the most prominent concepts and theories on that topic are briefly reviewed here in order to give an overview of the current state of the art. Even though most of the presented models originate from a context other than automated driving, they can certainly give an idea of what factors should be considered when assessing trust in this specific environment.

Trust model of Muir (1987, 1994) and Muir and Moray (1996)

Taking a look into the topic of trust in machines (specifically decision aid systems), Muir proposed a first model of trust in 1987. In her model, she describes trust on the basis of three dimensions of expectations (derived from Barber, 1983) and three levels of experience (related to the work of Rempel et al., 1985). Table 2.4 shows Muir's framework produced by crossing the two taxonomies of trust. Those dimensions are derived from interpersonal trust research, but due to the specificity and completeness of Barber's (1983) taxonomy, it is adopted as a basis for the development of a model of trust in machines by Muir (1987).

Table 2.4
Factors of trust (adapted from Muir, 1989)

	Basis of expectation at different levels of expertise		
Expectation	Predictability	Dependability	Faith
Persistence			
– Natural physical	Events conform to natural laws	Natural is lawful	Natural laws are constant
– Natural biological	Human life has survived	Human survival is lawful	Human life will survive
– Moral social	Humans and computers act "decently"	Humans and computers are "good" and "decent" by nature	Humans and computers will continue to be "good" and "decent" in the future
Technical competence	j's behavior is predictable	Has a dependable nature	j will continue to be dependable in the future
Fiduciary responsibility	j's behavior is consistently responsible	j has a responsible nature	j will continue to be responsible in the future

The dimensions of expectation include a dimension of *persistence*, which refers to the expectation that natural physical and biological as well as moral social orders are stable. The dimension of *technical competence* refers to the belief in the other agent to act in a predictable way. Finally, *fiduciary responsibility* refers to the expectation that the person to trust will act according to the interests of the other. These types of technical competence correspond to the taxonomy of behavior by Rasmussen (1983): everyday routine performance resembles skill-based behavior, technical facility can be interpreted as rule-based behavior, and expert knowledge refers to knowledge-based behavior. Those types of technical competence are crossed with the experience of a person on the levels *predictability*, *dependability*, and *faith*, implying that the dimensions of expectations and the dimensions of experience are orthogonal (Muir, 1994). As such, persistence, competence, and responsibility of the automated system are perceived by the human depending on his background experience with the automation (predictability, dependability, and faith). The person develops an expectation of the automation's characteristics and generates trust in the system. This trust can be more or less calibrated to the actual characteristics of a system. This way, the model proposed by Muir (1994) already includes considerations on balancing trust in an automated system dependence on the actual capabilities of the system. Muir (1994) suggests that while gaining experience with an automated system the nature of trust, which is first based on the consistency of the automation's behavior, will develop and become based on the perceived reliability of the automation. The highest level of trust would thus be achieved after prolonged experience with the system, when an operator can believe in the future dependability of the system. The basis of trust thus ranges from reason and fact to faith that goes beyond logical reflections (Adams et al., 2003). Later research and experiments conducted by the authors indicate that important aspects of trust in an automated system are captured by the interpersonal trust models taken into consideration (Muir & Moray, 1996). Using an industrial plant control task, people were asked to rate their subjective trust depending on the manipulated performance of the system. The results support the postulated model of trust, proving that trust was based mainly on perceived competence of the system. However, results also point to the importance of these factors depending on the time in trust development. For example, faith (as a rather emotional construct) has become apparent to be a better predictor of initial use of automation rather than of later stages of trust development (Muir & Moray, 1996). Based on this model of trust in a machine, Muir (1987) proposes several design guidelines that can help to design a decision aid system that is trustworthy (see Section 2.3.2).

Trust model of Lee and Moray (1992, 1994) and Lee and See (2004)

Similar to Muir (1989), Lee and Moray (1992) propose a relationship between different dimensions of trust formerly asserted by other research groups. Table 2.5 shows how they relate the dimensions to each other. While the propositions of Barber (1983) and Rempel et al. (1985) were included in the model proposed by Muir (1994), in the model of Lee and Moray (1992) the factors of trust are supplemented with their model representation. Lee and Moray (1992) suggest that the foundation of trust contains fundamental assumptions of nature and society. These assumptions allow the further layers of trust to develop (corresponding to Barber, 1983). The three constructs performance, process, and purpose are seen as the basic dimensions of trust (see also Wang, 2010). *Performance* is understood to be the current and former characteristics of an automated system, like its reliability, predictability, and ability. It relates to what the system is doing. The system is expected to perform in a consistent, stable, and desirable manner. *Process* is perceived as the appropriateness of the system's actions to manage a given situation. It describes how the system operates. This represents an understanding of underlying characteristics

Table 2.5
Factors of trust (adapted from Lee & Moray, 1992)

Barber (1983)	**Rempel, Holmes, and Zanna (1985)**	**Lee and Moray (1992)**
Persistence of natural laws	–	Foundation
Technically competent performance	Predictability	Performance (consistent, stable, etc.)
–	Dependability	Process (understanding behavior)
Fiduciary responsibility	Faith	Purpose (understanding intent)

of the system's behavior. The last dimension, *purpose*, is referring to the use case the system was developed for (Wang, 2010). It relates to why the system works the way the designer created it, thus describing the underlying intents of the system. In their supervisory control experiment, Lee and Moray (1992) report changes in human trust and control strategies during the interaction with an automated plant. Their analysis revealed effects of both system performance and system failures on subjective trust ratings, indicating that the factors influencing trust (performance and process) have an impact on other dimensions of trust (predictably, dependability, and faith) (Lee & Moray, 1992). The dimensions of interpersonal trust by Rempel et al. (1985) are thus suggested to be applicable to trust in automation as well. Lee and Moray (1992) furthermore found evidence that user's manual control abilities, next to trust, can influence system use.

Lee and See (2004) define the dimensions of detail and abstraction regarding the capabilities of the automation. The dimension *detail* refers to the specificity of trust (e.g., mode information or automation as a whole) and *abstraction* includes information about the performance, process, and purpose of the system (Lee & Moray, 1992). The authors recommend that both the level of detail and of abstraction should be respected when providing information to achieve highly calibrated and appropriate trust. Wang (2010) assumes that providing operators with information referring to these dimensions by training or interface design can lead to appropriate trust in a system. This assumption is examined more closely in Section 2.3.

Lee and See (2004) furthermore declare that information about the automation needs to be presented in consistency with cognitive processes that underlie the development of trust. In their research, they discovered that trust evolves through qualitatively different processes of information interpretation concerning the capabilities of an automated system. They differentiate between analytic-, analogical-, and affect-based comprehension of a situation. An analytic assessment of the situation, which includes a rational evaluation as a basis for further conclusions, can have an effect on trust. Likewise, an analogical approach based on category judgments evolving from direct experience or even indirect interaction with a system can mediate trust development. Finally, the most prominent aspect of trust is based on a rather emotional consideration of a situation: feelings and emotions of the user play an important role in the formation of trust (Lee & See, 2004). When these analytic, analogic, and affective processes of human information processing are considered in system design, this may also be reflected in balanced trust.

The elements described in the former section are parts of a larger process of trust formation suggested by Lee and See (2004). Figure 2.7 shows a conceptual model of how they envision the dynamic process that governs trust and its effect on reliance. Their trust model is one of the most commonly cited works on trust in automation. It describes the process of trust and reliance in detail, depending on a collection of individual, organizational, cultural, and environmental factors. The authors point out that a first belief formation already takes place based on the reputation of the system, gossip and observable interface features. Trust is formed based on this assimilation and dependent on a predisposition to trust as well as cultural and organizational influences. Depending on the current workload, perceived risk of system use, perceived self-efficacy of the human and other factors, trust results in an intention formation that finally leads to a reliance action. In this model, information about the automation is shown as one important factor for the human belief formation and resulting trust evolution. The level of detail necessarily plays an important role, varying from information about the system in general to detailed mode information. On an attributional abstraction level, ability, integrity, and benevolence can be considered as relevant factors (formerly described by Mayer et al., 1995). Other than that, similar distinctions are defined by Lee and Moray (1992) as performance, process, and purpose. Information through the display of the automation can support the appropriateness of trust regarding calibration, resolution, and temporal and functional specificity. However, it needs to be specified how this information could be provided to assure an appropriate development of trust in an automated system in a specific context like automated driving.

All in all, Lee and See (2004, p. 74) state that "appropriate trust and reliance depend on how well the capabilities of the automation are conveyed to the user". To support an appropriate development of trust, the different approaches to assimilate information regarding the automation (analytic-, analogical-, and affect-based) should be considered.

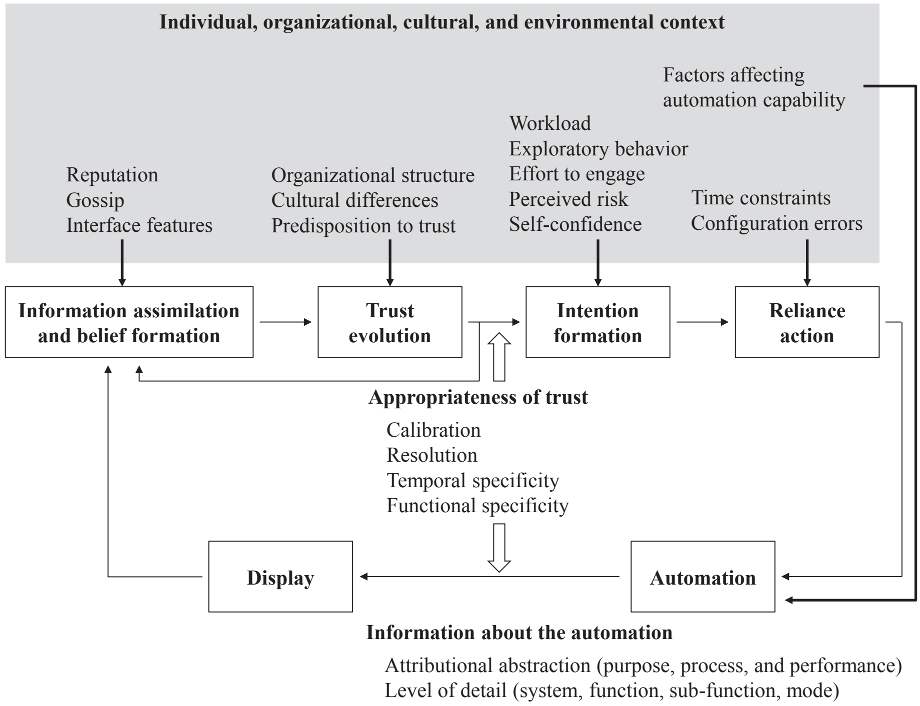


Figure 2.7. Trust model adapted from Lee and See (2004).

Trust model of Hoff and Bashir (2015)

A variety of different internal characteristics and experiences of the human as well as characteristics of the system, the situation, and the environment have shown to play an important role for developing trust in an automated system (Hancock et al., 2011; Lee & See, 2004; Merritt & Ilgen, 2008). A recent work of Hoff and Bashir (2015) has synthesized the current state of the art to a three-layered model of trust which addresses different aspects of trust characteristics as the main factors of trust development. It accumulates and synthesizes existing knowledge of trust development in automation. Furthermore, it distinguishes between layers that either become relevant in a specific context (situational trust, learned trust) or are seen as a permanent personal trait (dispositional trust).

Dispositional trust is seen as the overall tendency to trust in automation in general (not a specific system). It subsumes relatively stable individual factors, such as demographic and cultural aspects and personality traits of the user. For example, age, gender, and origin are known to influence the disposition to trust in a technical system. Personality characteristics and attitudes of a person have been proven to be even more important. Dispositional trust is seen by Merritt and Ilgen (2008) as trust in a system without any interaction with it. In the context of the current work, the individual demographics, personality traits, and attitudes mentioned in Section 2.2.3 are supposed to contribute to form a disposition to trust in automation that primarily determines trust in an automated vehicle upon encountering it for the first time. Situational and learned trust is dependent on the current situation: the external environment and context-dependent characteristics of the user play an important role, as well as past or current experiences of the interaction with an automated system (see Figure 2.8). *Situational trust* reflects the impact of external situational factors on trust. Most important external situational factors include task difficulty and system complexity. Environmental and situational factors can furthermore determine how much influence trust has on the actual reliance on the system. Reliance is guided by external factors like situational workload and perceived benefits and risks of using the automation. Considerations of Parasuraman and Riley (1997) underline the importance of these factors in dynamic allocation of function. *Learned trust* is based on the person's experience with a system. It represents a dynamic concept that forms over time based on a user's perception of the performance of the system. This perception goes along with a trust model of Merritt and Ilgen (2008), according to which trust evolves from dispositional trust in the beginning of the interaction to history-based trust due to further experience. Studies were able to verify that people's trust in systems adapts to the performance of the system with automation failures affecting trust considerably more than experiences with reliable function (Manzey et al., 2012; Merritt & Ilgen, 2008). Transparency of the system, on the other hand, has been proven to support trust in an automated system, even in the event of system failures (Verberne, Ham, & Midden, 2012; Ye & Johnson, 1995).

The model of Hoff and Bashir (2015) condenses other considerations and model representations of trust in automated systems. Trust is differentiated into several psychological constructs which allow for a structured analysis of trust regarding a variety of different antecedents and outcomes. The model is mainly based on research for automation in work environments, where trained personnel interacts with an automated system (e.g., aviation, military). These human-machine systems usually have to be used as part of the work task and the user often cannot freely decide to use or not use the automated system. The authors therefore recommend a transition of the model to more diverse automation that "people might encounter on a day-to-day basis" (Hoff & Bashir, 2015, p. 22).

2.3 Designing for trust in automated driving

Findings on trust development and trust manipulation can help to identify important aspects of designing a trustworthy automated system. This sections gives an overview over relevant research results on the influence of automation characteristics as well as on effects of human-automation interaction design. A collection of design guidelines is also presented.

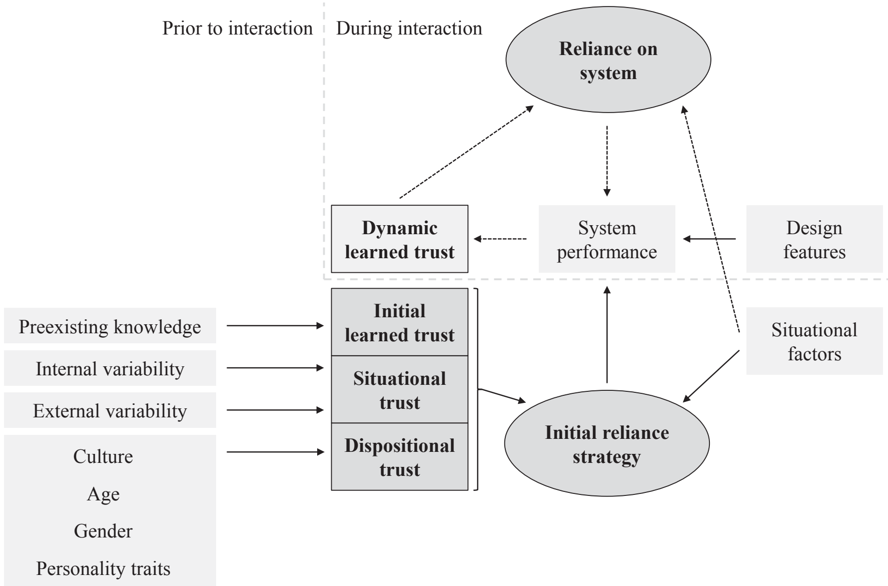


Figure 2.8. Model of operator trust adapted from Hoff and Bashir (2015).

Söllner and Leimeister (2011) have criticized that despite a large quantity of results gained in experiments regarding the development of trust, they find no translation of these insights into requirements for the design of technical systems. While research sometimes concentrates on the theoretical implications of experimental findings, it is crucial to find ways to include this research in the practical design of trustworthy systems (Söllner & Leimeister, 2013).

In Section 2.2.3, relevant factors influencing trust in an automated system were listed. Some of them (e.g., stable personality characteristics and attitudes of the human, former experience with automated systems, or situational circumstances) cannot be modified by the designer of a system—they need to be accepted as the basis on which an automated system is perceived. What can be altered by system designers is the human-automation interaction that takes place when using the automated system.

2.3.1 Findings on human-automation interaction

The characteristics of the automation that the user perceives are of paramount importance for the degree of trust. However good or bad the performance of a system may be, the crucial question is in what way the operator perceives the system's performance and how much his perception differs from the actual performance of the system. When designing for a flawless user interaction with a system, both the perceived performance of the system (level of automation, system reliability, system limits, and system failures) and the system's outward appearance (system transparency, predictability, and system complexity) need to be included in considerations. The aforementioned trust factors on the part of the automation (see Figure 2.6) are summed up in the following to point out research results of different domains that are helpful for the design of automated driving systems.

Perceived system performance

Without doubt, the performance characteristics of an automated system and its technical capabilities are crucial factors influencing the development of trust in a system. Researchers identified different parameters related to the performance of the system that matter in this context and that designers should bear in mind when creating an automated system.

Application of automation levels. Depending on how the user perceives the system's performance, he assesses how much of the task he can meaningfully allocate to the system. When the implemented *automation level* and the system's performance do not match, this can lead to difficulties in the interaction. Walliser (2011) found an influence of the automation level of an automated identification system on performance. The author attributed this effect to the improved trust calibration in a medium level of automation. Ruff, Narayanan, and Draper (2002) concluded from their experiments that when a higher automation level was used, even rare errors of an automation led to a significant drop in trust ratings. Their recommendation is a situation-specific application of automation levels to achieve optimal performance and trust. This recommendation is also supported by experiments that reveal overtrust in higher levels of automation that can lead to late or missing reactions to system errors (Niederée & Vollrath, 2009; Shen & Neyens, 2014). Also here, it can be concluded that the highest technically possible level of automation is not always the right choice. The perception of the system and the interaction of the user with it are decisive, and designers are advised to consider trust, but also workload and situation awareness when implementing a certain level of automation.

Clarification of system reliability and system limits. It seems logical that the better the *reliability* of an automated system and the fewer *limits* it has, the more trust a user will develop during the interaction with the system (see Section 2.2.3 Muir, 1994; Muir & Moray, 1996). It can, however, be extremely difficult (if not impossible) to create an error-free automation—after all, the system is built by humans, and humans can err. Together with the insights regarding the appropriate level of automation, it can be concluded that the level of automation should only be as high as the actual capability of the system allows. Otherwise, a high level of automation together with a high error rate and a low reliability can lead to misuse and distrust. If automation errors cannot be avoided completely, the level of automation needs to be made transparent. Research furthermore proved that the effect of performance and reliability of an automated system can be altered. When the system's boundaries are known in advance, trust is not necessarily affected by low system reliability, and a degradation of trust can be avoided (Adams et al., 2003; Beggiato & Krems, 2013; Dzindolet et al., 2003; Lewandowsky et al., 2000; Riley, 1996). This way, for example, recurrent smaller errors can be compensated (Lee & Moray, 1992; Ma, 2005; Muir & Moray, 1996).

Perceived system appearance

Already in 1991, Billings described that to create an optimal human-computer team, the interaction between those two partners needs to be designed in a certain way. When trying to alter or influence the formation of trust in an automated system in a certain way, many of the potential variables are related to an interface. A direct observation of the automated processes is often not possible, which is why a display is needed to mediate the perception of the automation-related information (Lee & See, 2004). Lee and See (2004) therefore suggest that the match between trust and the actual capabilities of the automation depends most of all on the two aspects *content* and *format* of a display. Content and format of the HMI are the adaptable parameters of trust in an automated system and could be "an important means of guiding appropriate expectations regarding the automation" (Lee & See, 2004, p. 73).

The so-called *Lens Model* by Llinas et al. (1998) that can be found in Figure 2.9 visualizes the idea that the system's appearance and its interfaces (x_K) are able to reflect the trustworthiness of the automation. An information transformation model originally introduced by Brunswik (1952) was used by Seong and Bisantz (1998) to create a model with the three components a) true state of the environment, b) observed state of the environment, and c) the operator's judgment based on his observations. The development of trust in this model depends on the observable characteristics of the system, namely the interface. The model stresses the importance of the operator's judgment of the automation, thus also addressing individual differences in development of trust in automation. The assumption that trust depends to a great extent on the interface of the system is an essential foundation of this work. Being able to promote a change in trust and guide the development of trust in an automated driving system is an important objective of the approach presented here.

Madsen and Gregor (2000) assume similar to Lee and See (2004) that trust consists of cognition- and affect-based processes. Based on the authors' understanding of trust development, affect-based trust is highly important in situations where the operator does not have enough information about the system to base his attitude on cognitive considerations. Transferred to the context of automated driving, it can thus be assumed that in order to support trust development, it is ad-

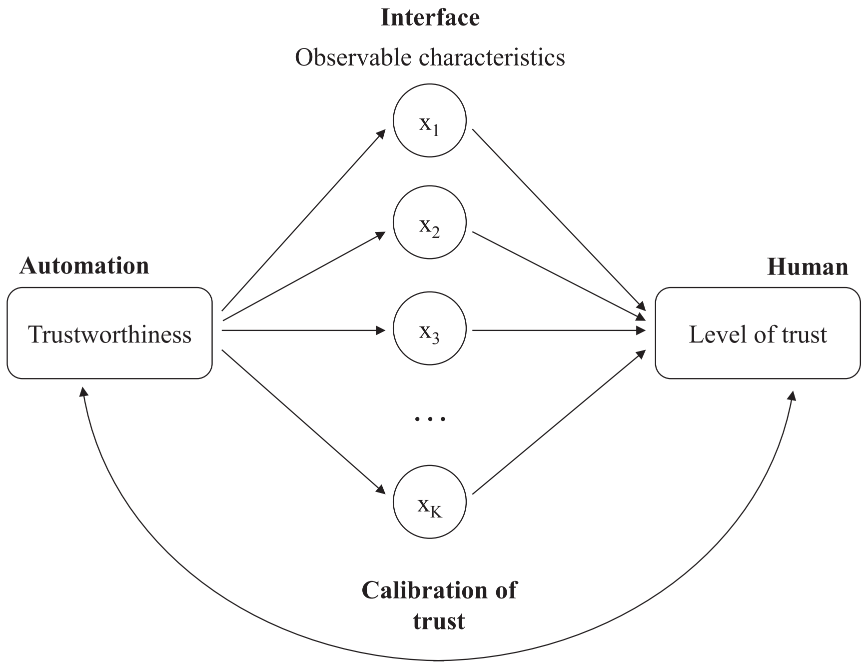


Figure 2.9. Lens model of trust (adapted from Llinas et al., 1998).

visable to provide enough information to the driver to avoid affective processes being the only basis to build trust on.

Not only the actual performance indicators and explanations can have an effect on trust—also the brand of the product, the perceived quality of the appearance, or the product design can greatly influence trust in a system. One design method is the design for etiquette (see Section 2.3.2). Another approach was taken by Waytz, Heafner, and Epley (2014), who were trying to strengthen trust in an automated driving system by giving the vehicle a name, a gender, and a human voice. They describe that in their simulator experiments, people who drove an anthropomorphized vehicle trusted their vehicle more. The drivers were also less stressed in an accident, and did not blame the vehicle or the system for an accident caused by another driver.

Reducing unpredictability and uncertainty. The dynamic situations in which human-machine interaction often takes place hinder operators from receiving the information they would normally need to manage a situation properly. They have to act under uncertainty, without having a profound knowledge about all the factors that might be relevant to the situation (Rajaonah et al., 2006, see Section 2.2.1). Considering trust as the confidence in another party under uncertainty (Lee & See, 2004), one could help the driver out of this dilemma by providing more information about the driving activities of the vehicle, thus reducing uncertainty to a minimum. This assumption goes along with findings of Ye and Johnson (1995). The authors distinguish between different explanations given by the system. *Trace* describes detailed record of reasoning steps, *justification* explains the logical argument, and *strategy* gives the higher-level approach. This gradation is easily comparable with the trust layers process (how?), performance (what?), and purpose (why?) introduced by Lee and Moray (1992). Results indicate that an explanation, especially justifications, can change the attitude toward the automation and make advice generated by an expert system more acceptable to users. These findings are expected to be applicable to domains in which decision making is highly consequential and the correctness of a decision is not easily verifiable (Ye & Johnson, 1995). Along with that, Verberne et al. (2012) found out that systems that take over a task of a human are judged more trustworthy and acceptable when they provide additional information rather than only fulfilling their task. In similar research, a group of drivers who were provided with uncertainty representation of an autonomous driving system took control of the car faster when needed, while they were, at the same time, the ones who spent more time looking at other things than on the road ahead compared to the control group without uncertainty information (Cai & Lin, 2010; Helldin et al., 2013). McGuirl and Sarter (2003) as well as Beller, Heesen, and Vollrath (2013) were able to show improved understanding of system and situation and better knowledge of system fallibility when confidence information of the system was provided, leading also to higher trust ratings and increased acceptance. Not only confidence information can influence the extent to which a system is perceived as trustworthy. Every information increasing transparency and supporting a more accurate mental model can help to create a trusted system, as research results presented in the next section show.

Enhancing transparency and mental models. Wang (2010) suggests that the components of a mental model can help to support appropriate trust in automation by providing an explanation for the system's behavior. According to Rouse and Morris (1985, p. 7), mental models are "the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states." Norman (1989) describes in his book "The design of everyday things", that the mental model of the human is developed through interaction with the system. The actual system image results from its physical structure, but it is not necessarily identical to the model the user has, even if the designer expects these images to be the same. The system image thus needs to convey the correct design model in a clear and consistent way, as this is the only way of communication between the designer and the user. The development of a mental representation depends on technical knowledge and other human characteristics, but is also contingent on the extent to which the automated system explains itself in a transparent way (Adams et al., 2003). The mental model is helpful for generating reasonable expectations about the automation. In that way, an adequate mental model can help calibrating trust to match expectations with the performance of the system (Beggiato & Krems, 2013; Itoh, 2012; Kazi et al., 2007; Ma, 2005).

A model described by Rouse, Cannon-Bowers, and Salas (1992) connects the mental model to the aforementioned factors of trust by Lee and Moray (1992) (see Figure 2.10). The first component is the descriptive function. It helps the person to gain knowledge of the system's physical description. The explanatory function relates to knowledge about system's operations and states. The last component, namely the predictive function, refers to the person's expectations about the system's future behavior and states.

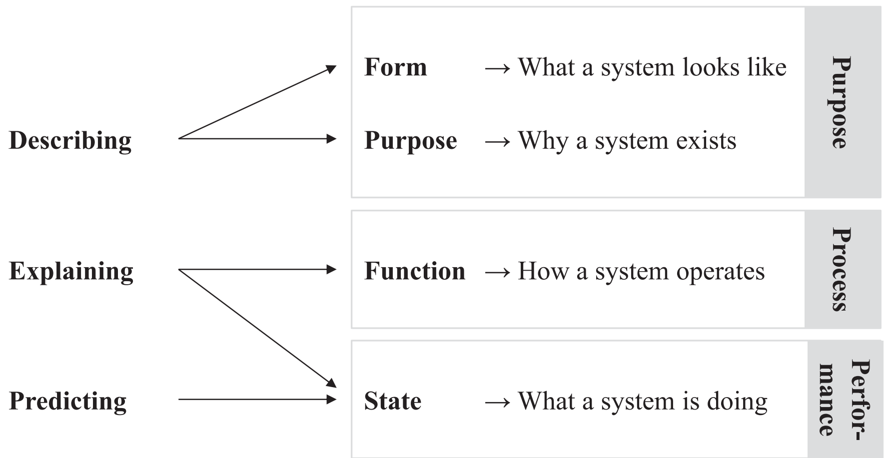


Figure 2.10. Nature of mental models (adapted from Rouse et al., 1992).

As Figure 2.10 shows, part of the development of a mental model is based on the transparent information of the systems state, function, and purpose. Especially the explanation of the system's behavior, the *explication of intention*, has been focused by research (Sheridan, 1988). The inner workings of the system need to be made clear to the user, and Adams et al. (2003) see this duty within the responsibility of the automation. This also includes the *explication of system limits*. With their experiment on transition ability from HAD, Merat (2014) show that people are better able to regain vehicle control when they are expecting automation to be switched off. Merat (2014) conclude that research needs to elicit more detail on how drivers can best be informed of their obligation to resume driving. Richards and Stedmon (2016) come to the conclusion that an optimal interaction between an automated vehicle and the driver can only be achieved when the system informs the driver of its actions and capability limits.

The findings mentioned above, along with many others (e.g., Adams et al., 2003; Beggiato et al., 2015; Beggiato & Krems, 2013; Itoh, 2012; McGuirl & Sarter, 2003; Simpson & Brander, 1995), advise to provide more information when using smart systems like automated driving systems. They suggest that giving information and thus providing a correct mental model of the system may lead to a higher level of system trust. Norman (1990) even goes to such lengths to say that the basis of the problem is not automation, but inappropriate feedback and interaction that cause failures in human-machine interaction. It can be concluded from the presented research that a system needs to provide transparent explanation to be understood and trusted. The mental model plays an important role in this relation as it enables the operator to predict the system's behavior as well as potential inaccuracies.

Reducing system complexity and the amount of information. Naturally, providing more and more information to the driver without paying attention to the limited human processing capacity does not lead to a relief of strain. Balancing the information provided and presenting them in a way the driver can actually assess and process is the fundamental challenge of human factors experts. Only then will the driver be able to adjust his trust according to the displayed information. To visualize this idea, Figure 2.11 shows a schematic illustration where the adequacy of trust is modulated by the amount of information given about an automated system (similar to the law of Yerkes and Dodson (1908) about the relationship between arousal and performance). It is assumed that trust is balanced best when an appropriate amount of information is given to understand the capabilities of the system without overloading the person with information they cannot process at the same time.

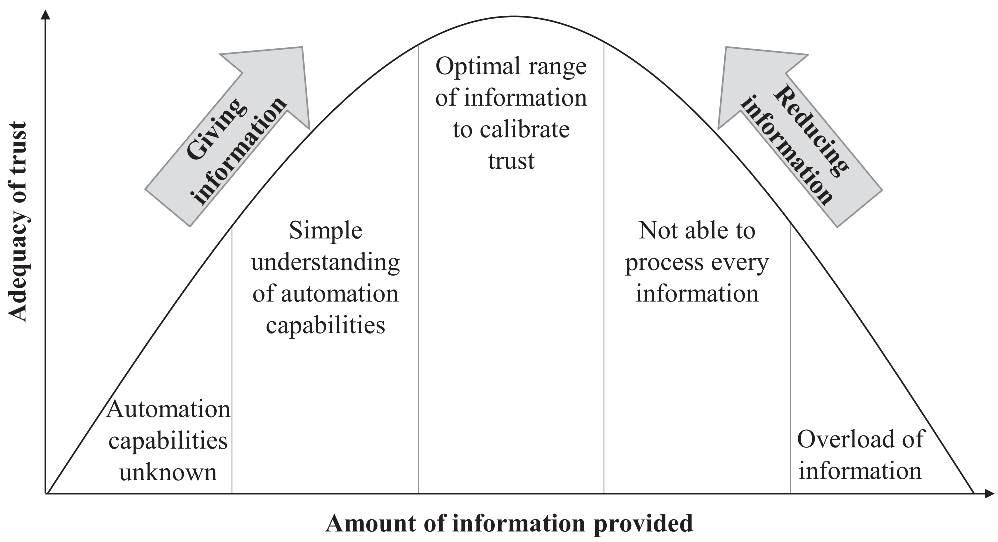


Figure 2.11. Trust modulation by the amount of information, similar to the law of Yerkes and Dodson (1908).

2.3.2 Design recommendations

The research presented above results in the question of how to design a trustworthy automated driving system in an appropriate way. The approach of designing a trustworthy system presupposes that trust is responsive to changes in the operator's perception of system properties (Muir & Moray, 1996). Trust is assumed to be subject to fluctuations and is considered a process rather than a stable concept (Atoyan et al., 2006), hence it should be possible to alter and possibly influence it.

Calibrating trust

In general, a correctly calibrated level of trust in automation should be strived for rather than the highest possible level. Lee and See (2004, p. 6) describe that "calibration refers to the correspondence between a person's trust in the automation and the automation's capabilities". This definition is closely linked to appropriate reliance, as trust can lead to a high or low use of automation and can thus result in an unjustified level of reliance (e.g., misuse or disuse, see Section 2.2.2). The optimum use of an automated system is thus achieved at a level of trust that matches its true properties. This good calibration is represented by the diagonal line in Figure 2.12, (Lee & See, 2004). The area above the line is characterized by overtrust, the area below by distrust.

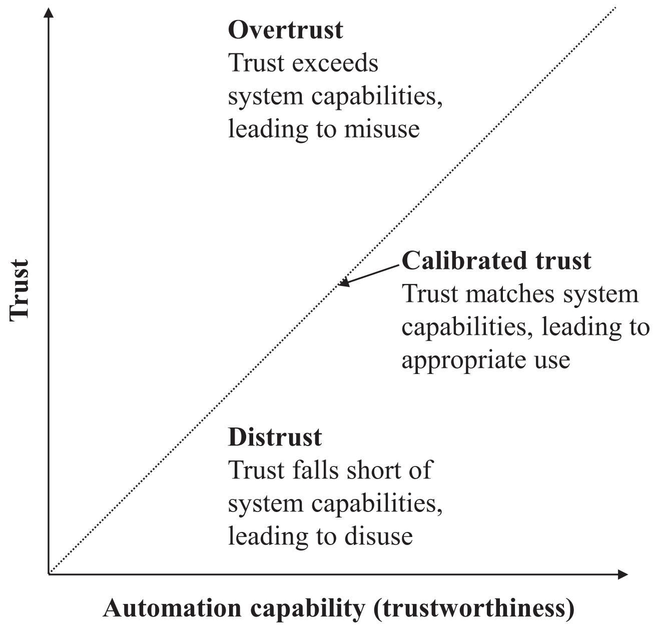


Figure 2.12. Trust and automation capability (adapted from Lee & See, 2004).

It is essential for automated systems that the level of trust matches its actual performance. However, for HAD systems a high level of trust is most appropriate, as the system has a high level of capability. Calibrated trust in a HAD system mostly means that the driver's confidence in the system is high enough to relax (see Helldin et al., 2013). Per definition, no errors of the automation are expected in this level of automation, and system boundaries can be announced in advance. The only relevant situation may be the takeover situation: trust should not be excessively high, ending in a potential lack of reaction to a takeover request. The calibration of trust is not a focus of this work due to the limitation on a high automation level with high capabilities of the system. Nonetheless, research results on trust calibration may help in designing a trustworthy automated system.

To achieve correctly calibrated trust, high resolution and high specificity of trust are required according to the model of Lee and See (2004). Then, the range of trust relates to the range of varying system capabilities and the showed trust differentiates between specific modes, always adjusting to the appropriate level (Popken, 2009). In summary, as Lee and See (2004, p. 73) put it: "If the information is not available in the display or if it is formatted improperly, trust may not develop appropriately.". To (re)calibrate trust, Muir (1987) recommends several methods, most importantly improving the accuracy of operators' perceptions of machine competence (Muir & Moray, 1996).

Clearly, the subjective impression of a system can influence trust. The match between expectations and the actual capabilities of a system is relevant to the increase or decrease of trust. Ways to increase this match between the operator's expectations and the real performance of the system are thus related to the enhancement of knowledge about the system. *Advance knowledge of system boundaries and potential failures* can reduce uncertainty in the interaction with automation. This is underlined by results of Riley (1996), who reports that knowing the shortcomings of the automated system helps the operator to maintain their level of trust even in the event of a failure (see also Beggiato & Krems, 2013; Dzindolet et al., 2003). *Transparency and salience of system actions* in general is assumed to be able to enhance trust to an appropriate level, as has been shown in the previous section. In short, "the extent to which the system can be predicted is as important or more important than the extent to which the system is reliable" (Adams et al., 2003, p. 33). To raise trust in a HAD system to an appropriate level, interaction design recommendations are collected in the next section.

Design guidelines

To design usable and acceptable systems, many authors provide guidelines to follow (Billings, 1991; Christoffersen & Woods, 2002; Herczeg, 2014; Kaufmann, Risser, Geven, & Sefelin, 2008). However, only few authors actually address design rules for promoting the trustworthiness of an automated system. Approaches relevant for this work are highlighted in the following.

One method is the implementation of *adaptive automation* with varying levels of assistance by the system. This way, the operator stays in the loop and still has at least partial control of the task on hand (Dijksterhuis, Stuiver, Mulder, Brookhuis, & de Waard, 2012; Kaber & Endsley, 1997; Miller & Parasuraman, 2007; Parasuraman & Riley, 1997). Of course, if fully automated driving shall be achieved, regular disengagements of the automated system are not practical, and shared control cannot be a long-term solution. Merat (2014) instead recommends considering how to remind drivers of their obligation to resume control.

Another approach is the *etiquette-based design* of automation suggested by Lee and See (2004) and Miller (2005) that takes the suggested analytic, analogic, and affective parts of trust into account. According to them, enhancement of the interaction could be achieved by recognizing the influence of social context and designing an automation to have a socially acceptable behavior (for example, by using colloquial language of the domain). Miller (2005, p. 4) understands etiquette as “the largely unwritten codes that define roles and acceptable or unacceptable behaviors or interaction moves of each participant in a common ‘social’ setting”. Their experiments verify that a good etiquette can even compensate for a low reliability of the automation, at least during a long-term relationship. Also, Spain and Madhavan (2009) discovered that when using an imperfect automated aid, a polite system is perceived as more reliable and trustworthy than an aid without etiquette. This result points at the importance of interface features and the human-machine relation in comparison to the actual capability of a system.

A promising approach to trust tuning that describes the main approach of this work is the careful design of the *system’s interface*. Trust tuning can take place when the interface helps the human operator to adapt his reliance on the system based on the system’s capabilities. This goes along with findings of Seppelt and Lee (2007). Their results suggest that informing drivers continuously about the automation state can be more effective than only warning the driver in case of an emergency. Cai and Lin (2010) summarizes that a well-designed interface can support the transition of control between a human and an automation. The idea to use the system’s interface to design for trust in automation was followed up by other researchers, and design guidelines were developed specifically for promoting trust. An overview over the recommendations is given and summarized at the end of this section.

Design recommendations by Muir (1987, 1994) and Muir and Moray (1996). As stressed above, the communication and transparency of the system are crucial when designing trustworthy automation. Already in 1987, Muir described how calibration of trust in an automated system could be improved. As assumed by Muir (1994), trust develops through a learning process and should thus be modifiable by training the operators. For designing decision aid systems, Muir (1987) recommends *improving the user's ability to perceive a decision aid's trustworthiness.* This could be achieved by increasing the observability of system behavior as well as the transparency of the automated function. This way, the user is provided with evidence that can be compared with the information given from the system. Also, to *modify the user's criterion of trustworthiness*, Muir (1987) advises to make the system's expertise, capabilities, responsibilities, and boundaries explicit. A criterion level of reasonable performance could be provided, as well as a clear comparison of overall performance when using or not using the automated system. It is furthermore recommended to *enhance the user's ability to allocate functions in a system*. By assigning the human to be responsible for dynamically allocating functions to the system, he still has the responsibility for decision making and is thus not alienated from the automation. Lastly, the guidelines for designing automated systems suggest *identifying and selectively recalibrating the user on the dimensions of trust which are poorly calibrated.* Reasonable expectations towards the automation should be specified via training. The source of badly balanced trust can then be identified and improved selectively.

Design recommendations by Atoyan et al. (2006). According to the findings described above, Atoyan et al. (2006) also developed guidelines for appropriate system design with particular emphasis on promoting trust in the system. Their general design rules are derived from a review of theoretical, empirical, and experimental studies. The authors recommend designing a system for an appropriate level of trust that is neither too high nor too low. Overtrust and undertrust are both considered to undermine system safety and profitability, and should thus be avoided. Also, both the system and the user should be prepared for system boundaries, especially during the introduction of a new system. The impact of initial experience on trust is emphasized, because a system that is not trusted in the beginning will not be used at all and trust can never be developed. Regarding the interface, Atoyan et al. (2006) suggest organizing the information according to user expectations. User-centered design could help to implement human-machine interaction according to the user's expectations. Lastly, cultural and individual differences should be considered, as they have been shown to influence expectations and development of trust in automation. They should thus be addressed by appropriate training. Concrete guidelines are also provided in an attempt to help designers of automated systems (or specifically decision aid systems) to support appropriate trust tuning. Having tested those design guidelines, Atoyan et al. (2006) were able to confirm an increase in trust due to enhanced usability of the user interface. Next to other interface qualities, informative feedback and guidance have been found to be of major importance.

Design recommendations by Hoff and Bashir (2015). Hoff and Bashir (2015) collected results from diverse research papers and summarized the implications for designing automation in five guidelines. They recommend paying attention to the *appearance and anthropomorphism* of the automated system. Anthropomorphism is understood as a process whereby people attribute human characteristics like the capacity for rational thought and conscious feeling to nonhumans Waytz et al. (2014). Increasing anthropomorphism can promote greater trust, but other factors must be taken into account. Age, gender, culture, and personality of potential users need to be considered because the design may impact their trust differently. Also, the *ease of use* should be promoted by simplifying the interfaces and increasing the saliency of automation feedback to promote greater trust. When it comes to the *communication style*, Hoff and Bashir (2015) suggest ensuring an adequate appearance and increasing the politeness of the automated system. Along with other researchers, Hoff and Bashir (2015) stress the value of *transparency and feedback* of the system. Their guidelines recommend to provide users with ongoing feedback concerning the reliability of the automated system, depending also on situational factors, and to communicate explanations for automation boundaries or even failures. Lastly, user preferences need to be considered for the appropriate level of human control during system interaction.

While some research addresses automated control systems and decision support systems in general (Atoyan et al., 2006; Hoff & Bashir, 2015; Muir, 1987), others belong to a specific domain like aviation (Miller, 2005; Spain & Madhavan, 2009), industrial production (Lee & Moray, 1992; Muir & Moray, 1996), or military (Dzindolet et al., 2003; Rovira, Cross, Leitch, & Bonaceto, 2014). Only a few recommendations explicitly target the automotive area (Beggiato et al., 2015). To sum up the design recommendations appearing in literature, Table 2.6 presents the design guidelines considered most relevant for the design of an automated driving system. The work at hand makes use of the design recommendations introduced in this chapter and tries to implement them in a design concept for HAD vehicles (described in Section 3.3). The HMI concept designed for this research focuses on the enhancement of the automated driving system's transparency, and is utilized to evaluate a model of trust in automated driving that is presented in the following chapter.

Table 2.6
Design guidelines relevant for creating trustworthy automated driving systems, summarized from Muir (1987), Atoyan et al. (2006), Hoff and Bashir (2015) and others

Design Guideline	Explanation	Source
Simplify the interface	Make automation easy to use.	Hoff and Bashir (2015)
Provide access to raw data	When raw data is still available, low reliability of a system might not be followed by serious consequences because the human can intervene at any time.	Rovira et al. (2014)
Make the purpose of the automation clear	The purpose of a system is, next to the factors performance and process, an important detail for the development of a mental model of the system.	Atoyan et al. (2006); Lee and Moray (1992)
Design with good computer etiquette	A good etiquette of the system will leave a positive impression and can influence the development of affective and analogical trust positively. Thus it is recommended to increase politeness and anthropomorphism of the system's communication style. Of course it needs to be stressed that good automation etiquette should not be used to compensate for insufficient system reliability.	Hoff and Bashir (2015); Miller (2005); Spain and Madhavan (2009)
Reveal the rules and algorithms used by the automation	This guideline refers to the explication of the trust factor process. When the operator has the possibility to track important system decisions, he can understand the system better and will trust it more.	Atoyan et al. (2006); Lee and Moray (1992)
Provide the user with ongoing feedback	Feedback should be given concerning the reliability of the automated system and the situational factors that can affect its reliability. The user needs to be aware of dependencies between system and environment to not be surprised by changes in the system's behavior due to varying context. If the system's performance is context dependent, the context should be made explicit to the user.	Hoff and Bashir (2015); Muir (1987)
Provide means to indicate unreliable data	Missing, incomplete, or invalid data needs to be recognizable. As has been shown before, predictability of system boundaries can diminish their effect on trust and low reliability can better be compensated for. Whenever possible, the distinction between functions of differing reliability should be made clear to allow for an independent evaluate of these functions.	Beggiato et al. (2015); Dzindolet et al. (2003); Lee and Moray (1992, 1994); Muir and Moray (1996); Riley (1996)
Show the source of low automation reliability	As was explained in detail in Chapter 2.2.3, trust decreases as a consequence low reliability or the experience of system boundaries. This can be prevented when explaining the system boundary to the user.	Beggiato et al. (2015); Dzindolet et al. (2003); Riley (1996)
Train the operator	Research in aviation domain shows us that trained operators are less prone to effects like automation bias and complacency.	Atoyan et al. (2006)

2.4 Summary and conclusions

In the beginning of this chapter, automated driving was discussed as a new opportunity of modern transport. As a first step, the underlying definition of this particular kind of automation was outlined. The focus of this work lies on trust in conditional and highly automated driving (levels 3 and 4 of the SAE taxonomy) in the context of highway driving. While automated systems have numerous advantages, the different levels also inherit automation effects to a varying degree. The possible downsides were discussed, as well as challenges that still need to be overcome before the new technology can be brought on the street.

One of the upcoming tasks to be solved is the promotion of trust in the novel technology. The thematic focus of trust in an automated driving system was inspected in more detail in this chapter. This psychological factor gains importance as drivers are more and more obliged to hand control over to the vehicle. In this work, trust is understood as the attitude that an automated system will act according to the human's objectives in an uncertain situation. It can result in an over- or underreliance on an automated system. A diversity of human and machine characteristics play a role in the process of trust and reliance intention. It was demonstrated that trust in automated driving is still a young area of research, but can build upon and take up a diversity of findings and models developed in former research of similar domains.

While former models and research results on trust in automation were mostly derived from studies on decision aid or advisory systems, when developing HAD functions, trust similarly needs to be considered in this novel context. The presented research models jointly stress the importance of an appropriate communication between the automated system and the human user. To address this need for transparency, the work at hand focuses on the development and manipulation of trust in HAD, with a particular emphasis on prospective potentials of an HMI concept. To this end, this chapter provided background on research on human and machine characteristics in conjunction with system transparency. Also, design guidelines were introduced that can help to create transparent automated driving systems.

This dissertation aims at working out a specific HMI concept for trust in automated driving. A comprehensive model of trust in this new technology serves to develop an understanding of what factors are relevant to alter trust in the technology. Chapter 3 discusses this endeavor in more detail and presents an applicable working model of trust specifically for the context of automated driving. Open research questions are formulated and HMI concepts are developed to be tested in the user studies.

3 Research concept

This chapter introduces a comprehensive working model of trust in automated driving based on the models of Lee and See (2004) and Hoff and Bashir (2015), which forms the basis for the following research (Section 3.1). Section 3.2 describes the open research questions that are addressed in the studies presented in Chapter 4. The last section of the chapter finally describes a design approach developed to investigate trust in an automated driving system (Section 3.3). Section 3.3.1 outlines first exploratory studies conducted to gain indications for the design and implementations of the main studies. To conclude the chapter, the HMI concept for automated driving as it was used in the test vehicle is presented in Section 3.3.2).

3.1 Proposed model of trust in automated driving

As has been made clear, a diversity of factors can influence trust development in automated driving. Combining the research on trust in automated systems and the introduced models of Lee and See (2004) and Hoff and Bashir (2015), Figure 3.1 illustrates the concept of trust that is used in the work at hand. It is especially focusing on aspects relevant for the context of automated driving. Related to the factor overview of Hancock et al. (2011), main factors that research found to be relevant for trust in HAD are depicted. Those are either human- or system-related or describe a certain aspect of the environmental situation. The human-related factors, on the one hand, are seen as the basis for dispositional trust. This trust depends on the person's characteristics of personality, e.g., traits, attitudes, states as well as on ability-based characteristics, e.g., system experience. As reported, some research also found demographic factors like age to play an important role for dispositional trust. Learned trust, on the other hand, is developing based on the experience of system characteristics. These can be performance conditions, e.g., system reliability, system behavior, but also aspects of system design and human machine interaction, e.g., transparency and appearance of the system. This differentiation between dispositional and learned (history-based) trust had been suggested by Merritt and Ilgen (2008). The user needs to find a balance between his readiness to trust and his perceived adequacy to trust based on an assessment of the system's skills. The outcome of the resulting overall trust in the automated driving system can be observed as trusting behavior, such as an allocation of control

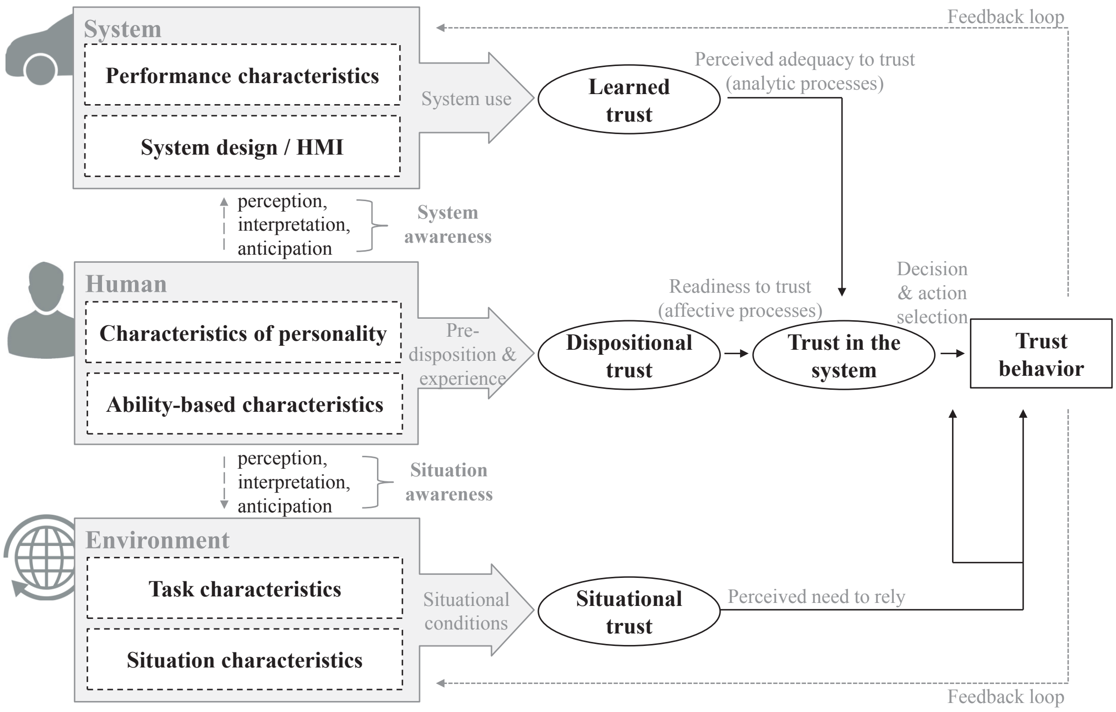


Figure 3.1. Suggested working model of trust in HAD, based on the models of Lee and See (2004) as well as Hoff and Bashir (2015). It depicts relevant factors influencing trust in an automated driving system and the resulting trust behavior.

to the automation, reduction of monitoring behavior, and an increased orientation toward a non-driving-related activity (Popken, 2009). Situational trust depends on environmental factors, like characteristics of the current situation and the task at hand, and influences resulting trust behavior. This perceived need to rely can be understood as a behavioral adaptation to the surroundings based on the underlying demands. It rather affects the influence trust has on reliance (decision and action selection).

The factors mentioned above are considered important variables influencing the perception of and interaction with an automated driving system. Of course, the possibilities to influence the characteristics of the human user are limited. States and ability-based characteristics may be influenced to a certain degree, e.g., by informing the driver about the system and giving him the possibility to try the system out. Personality traits, however, are invariant and stable. Also, characteristics of the environment are difficult to alter, even though they are variable. Some driving situations can be avoided or their criticality can be reduced, e.g., by reducing driving speed. However, some characteristics of the driving situation are given and cannot be changed, e.g., weather conditions. The most promising adjustments can be made on the part of the system itself. On the one hand, the performance of the system is a crucial property: the level of automation, its reliability, dependability, and behavior have already been shown to be of high importance for system and trust evaluation. On the other hand, the system's appearance and its transparency, which is communicated through the interface, are of special relevance.

Hoff and Bashir (2015) recommended a verification of their model with regards to automation that may be encountered in everyday life. In this work, the model is used for the specific context of automated driving. In this context the human is not trained to use the automation and may therefore have to be supported differently. The aspects of the working model presented here condense the considerations of Lee and See (2004) and Hoff and Bashir (2015) while paying particular attention to HMI design as a relevant regulating unit for trust in automated vehicles.

As has been shown in this chapter, designing for transparency of a system can support a better understanding and a correct mental model of the system. This can in turn promote trust in a system. When the expectations regarding the system match its actual capabilities, trust is most likely to arise. It is assumed that through interface design, it is possible to influence trust in such a system and regulate trust related behavior. Through that, the use of automated driving systems shall be made as safe and comfortable as possible for the driver.

3.2 Research questions

Trust in automation is one of the major predictors of the intention to use a system. It can possibly be influenced by accordant interaction strategies and HMI concepts. This work concentrates on the evaluation of such a concept for an automated vehicle regarding its effect on trust development and trust maintenance. An elaborate literature analysis was undertaken to give insight in the current state of research and enhance the understanding of the psychological constructs addressed in this work. With this, a methodical approach is pursued to evaluate how trust in automated vehicles can be enhanced to an appropriate level to relieve the driver of the strain of driving. As stated before, trust is expected to be achieved by providing information about the system that improves the driver's understanding of its functionality. More research is needed to identify indications that should be used to convey precise information during automated driving. To address this need of research, several user studies are employed to answer the following research questions.

What impacting factors and correlates for trust exist regarding the interaction with an automated driving system? As a first step, it shall be clarified whether certain dispositions of the driver, aspects of the system or of the interaction between driver and system have a measurable effect on the level of trust in an automated driving system in particular. The driver might have certain dispositions that guide his initial trust in an automated driving system. As research showed, some personality traits have a significant impact on trust in automated systems in general. This work shall find out if this is true for automated driving systems as well. Regarding the aspects of the system or system interaction, it shall be found out which interaction strategy is able to guide system trust and establish an acceptable level of trust in a formerly unknown automated driving system. All characteristics that are taken into account are derived from the research described in Section 2.2.3. It shall be evaluated how trust in automated vehicles evolves and how this development can be regulated. A correct mental model (compared to the real model of the system) is needed to ensure the effective functioning of the human-machine interaction. Research results suggest that operators can establish calibrated trust in an automated system better if they have an appropriate mental model of the system (Wang, 2010). A mental model consequently provides a basis for appropriate trust in an automated driving system. It cannot always be assured that users of a system are familiar with all its capabilities and processes beforehand. Thus, the system itself should be able to provide information. To give the driver the possibility to develop a correct mental model of a system and establish trust in it, an HMI concept giving feedback about the systems state and behavior can be integrated into the system. The HMI concept should give descriptive information (e.g., the functionalities and competencies of the automated driving system), explanatory information (e.g., boundaries of the automated driving system, reasons for takeover or system limits), and predictive information (in the context of automated driving, e.g., the detection of surroundings, upcoming actions and maneuvers, or the predicted time until takeover) to convey a correct mental model to the user (Rouse et al., 1992).

Can system transparency engender a pertinent level of trust in an automated driving system, even in the event of a system limit? Trust is considered a relatively stable construct, but it is nonetheless subject to changes. It is altered especially through system interaction and experience with the system, system performance being one of the major factors influencing trust. Thus the effects of experiencing a system limit or even a failure have always been in the focus of automation research. When it comes to the context of automated driving, the alteration of trust consecutive to system limits or handovers is of major interest. As stated before, it is assumed that giving more information about the system's behavior but also about boundaries can enhance trust in the system. Several results of experimental studies hint in that direction (e.g., Adams et al., 2003; Riley, 1996). Thus, it shall be clarified in which situations an information is of outstanding importance, and in which situations the driver has either no capacity or no interest in receiving further information about the automated driving system. This way, the user should not be overwhelmed by all the information technically possible, but should receive all the information necessary to maintain his level of trust in the system.

How does trust in an automated driving system evolve and what are the connotations of trust in different stages of system use? It is furthermore assumed that once trust is established, less information is required to maintain the mental model, as the concept is considered to be fairly constant (Parasuraman et al., 2008). Numan (1998) expects trust to increase in the long term. He describes that once trust has been established it can be altered temporarily, but never completely disappears (Rajaonah et al., 2006). Going along with that, it is assumed that the established trust in the automated driving system is consistent rather than fleeting, even after taking the information away.

The research questions are summed up in Figure 3.2. They are associated with the studies presented in Chapter 4 that are designed to answer the questions. All studies furthermore aim at identifying subjective and objective measures for trust in automated driving.

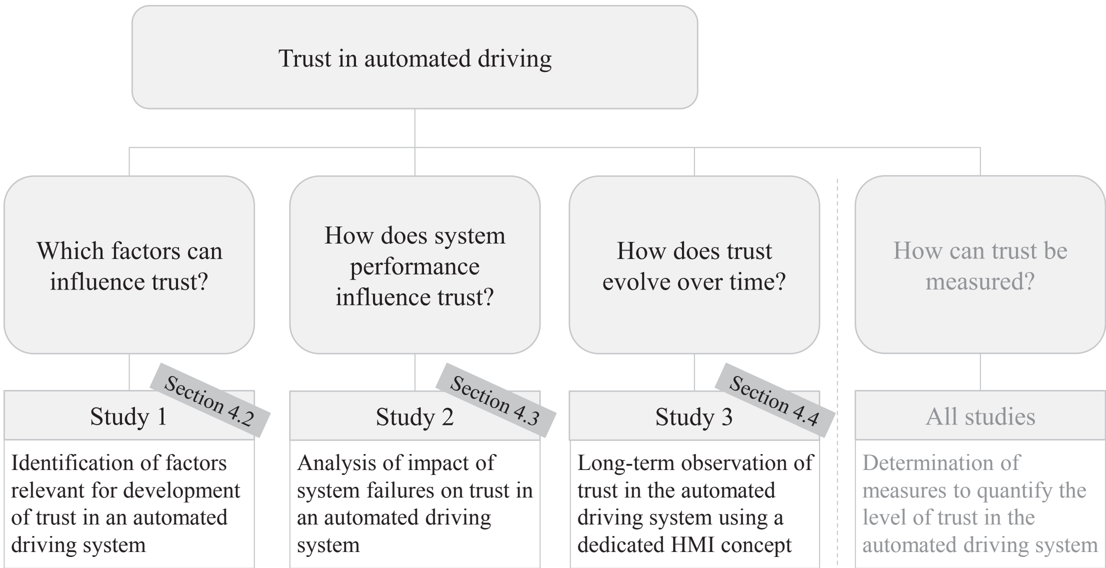


Figure 3.2. Research questions and corresponding user studies.

3.3 HMI design

Before attempting to answer the proposed research questions in detail, three exploratory user studies were conducted. The objective of these investigations was to develop an idea of how a user-centered HMI concept for automated driving could look like. They are presented in Section 3.3.1 before introducing the final HMI concept used for the main studies in Section 3.3.2.

3.3.1 Insights from initial studies

To develop an HMI concept on the basis of the design recommendations for automated driving (presented in Section 2.3.2), first prototypes were tested in small and qualitative study settings and comments were collected to improve the concept. The findings and insights of these were integrated in the design of the main studies, which are described in detail in Chapter 4.

Pre-study 1

No naturalistic driving studies in an actual automated vehicle were known at the time this research started. Thus, to explore how people react when sitting in a self-driving vehicle in real traffic conditions for the first time in their life, a small test run was conducted to guide expectations for the following studies.

Participants. To examine how users react when sitting in a self-driving car for the first time, a test drive was set up, giving a small sample of 20 drivers (30.35 years on average, $SD = 6.52$ *years*; 4 women, 16 men) the opportunity to be driven by a prototype self-driving vehicle. Participants were all employees of the Electronics Research Laboratory (Belmont, California), a research laboratory that is part of the Volkswagen Group of America. They all were first-time users of self-driving cars who did not know the system beforehand.

Method. The setup within the cockpit is illustrated in Figure 3.3. For this test run, a sparse HMI concept was used, displaying instructions and a countdown until takeover in the instrument cluster (1), the current mode in an LED bar below the windshield (2), and the status of the automated driving system in a small center console display (3). More details regarding the prototypical vehicle can be found in Section 3.3.2.

The route of the test drive led participants on different urban and city roads around the Electronics Research Laboratory. The drive lasted 45 minutes, with approximately 35 mph maximum speed. On some roads, the automated driving system was ready to take control, others needed to be driven in manual mode. When approaching one of the four sections of automated driving, the drivers received an indication that they could now engage in automated driving. Each participant was ideally presented with eight transitions between manual and automated driving per trial. An indication was composed of a distinct tone, enhanced peripheral LED lights (Figure 3.3, Element 2), and an information in the instrument cluster (Figure 3.3, Element 1). Having handed over control by pressing two buttons on the steering wheel, participants could experience how the car steered, stopped for red traffic lights or other cars, and drove on when the traffic situation allowed it. The active status of the automated driving system was always visible for the

Figure 3.3. Setup of the first pre-study with HMI elements for HAD (1 – instrument cluster, 2 – LED bar, 3 – center console display).

driver in the small display of the center console (Figure 3.3, Element 3). Furthermore, drivers could engage in infotainment activities—such as surfing the web or reading e-mails—using the infotainment features in the large center console display. These features were available until the automated driving system announced an upcoming takeover and drivers had to take the wheel again.

Questionnaire. The survey of the pre-study consisted of three parts. Before the test drive, general personal information like gender, age, and driving expertise were collected with an online questionnaire (similar to Appendix A.1.3, Table A.6 and Table A.7). This questionnaire also contained several open questions about participants' current thoughts about automated vehicles, expected advantages and disadvantages of automated vehicles, and expectations regarding the upcoming test drive. During the test drive, a qualitative interview was used to collect the overall acceptance of the HMI features and suggestions for improvements of the interface design. Also, the stress level during the different parts of the route was queried (see Appendix A.1.3, Table A.8). After the test drive, participants were asked to fill out a self-designed online questionnaire to give specific feedback on the HMI elements and takeover indications used in the vehicle.

Results and discussion. Participants liked the idea of automated driving and considered it attractive. However, most people did not trust the self-driving vehicle completely and felt the need to supervise the behavior and actions of the car. The prototype's performance was considered acceptable and people liked the HMI concept of the prototype with LED bar, sound, and instrument cluster. They considered the LED bar most useful for mode information, whereas the instrument cluster and the sound were seen as most useful for transitions back to manual mode. More importantly, though, drivers wanted more information about the drive. Beyond the status information given in the small center console display, they demanded to receive more feedback about the intended driving behavior, e.g., upcoming maneuvers like lane changes, turns. Also, route information was important for the drivers, e.g., on a map displaying automated driving sections of the trip. Finally, a representation of the vehicle's detection was desired, e.g., other cars, traffic lights, bicyclists, and how it interprets these surroundings. When this information is provided, participants of the pre-study can imagine gaining enough trust in the system to relinquish the driving task. The interviews during and after the test drive furthermore revealed that 70% of the drivers felt more stressed during the automated drive compared to the normal manual drive. They felt out of control, were unsure about the capabilities of the system, and needed to gain trust in the system as they were not used to it yet. Also, the driving style in automated mode seemed to play an important role. It was suggested to adjust it depending on the particular driver (e.g., personalized) to match the individual driver's style.

Summary. This first investigation of trust in an automated driving vehicle provided insight into the need for information the driver has during an automated drive. It can be concluded from the interviews and questionnaires that even though people are positively inclined towards automated driving technology in general, they want to be informed about the drive when they are no longer in control (see also Beggiato et al., 2015). The goal of the next pre-study was to find out more about the preferred way of interacting with the system and the needed information about the automated driving system.

Pre-study 2

As a follow-up to the first test drive, first thoughts were spent on the design of a more detailed HMI concept. As people were nervous about transferring control to the vehicle, it seemed necessary to guide people to use such a system without fear. Losing control of the vehicle led to the request of being informed better about the system. In a simulator study, a first concept was tested to confirm this idea.

Participants. 40 participants who were not familiar with the HMI concept for automated driving were invited to drive in a mock-up with a simulation of a highway on a large screen in front of them. The participants were on average 29.23 years old ($SD = 5.09$ *years*; 7 women, 33 men). Participants took part voluntarily and were employees of the Electronics Research Laboratory, were the study was also conducted.

Method. A small driving simulation was used as an experimental method to evaluate the design of the new cockpit concept. The driving experience lasted for approximately 15 minutes. Two different highway scenarios were driven in the simulated environment. One situation consisted of a slow car ahead and the other one included a traffic jam. The order of the two scenarios was randomized to avoid learning effects. Participants started driving in manual mode, until the automated driving function was offered and people could engage the system. One group of participants activated the system implicitly by letting go of the steering wheel, the other group could use a button in the middle console to explicitly activate the system. The steering wheel retracted in automated mode to give way for participants to engage in non-driving-related activities. Participants had the possibility to read web pages or watch video clips located in the instrument cluster and controlled by touch pads on the steering wheel when automated mode was activated. The setup is shown in Figure 3.4.

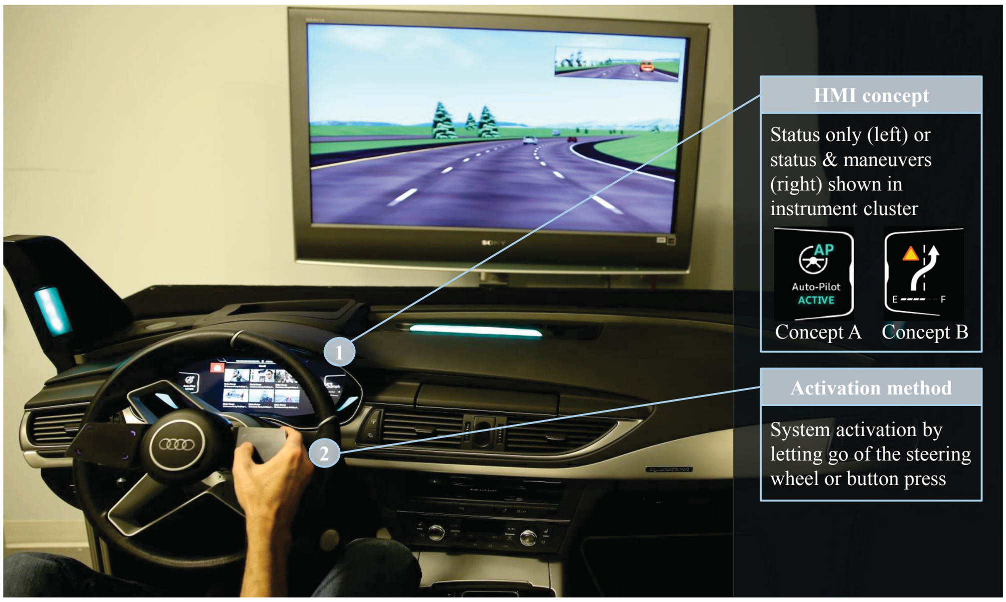


Figure 3.4. Setup of the second pre-study with the HMI concept and infotainment functionality located in the instrument cluster (1). The system is activated by letting go of the steering wheel or a button press (2).

Next to the two activation methods, two different HMI concepts were available to be compared in the study. Half of the participants saw version A, showing necessary information about the system mode. Version B also included indications showing the driver what the automated driving system is doing, e.g., maneuvers, acceleration, and traffic information. Example representations of the concepts can be seen on the right side of Figure 3.4.

Questionnaire. Personal data and driving experience where collected in the beginning of the study to describe the sample (see Appendix A.1.3, Table A.6 and Table A.7). Participants were asked about their trust in the automated driving system with the help of a questionnaire by Madsen and Gregor (2000) (Appendix A.1.3, Table A.4). After the simulated drive, participants were asked how they felt during the drive and about the specific indications they saw during automated driving.

Results and discussion. There was no preference in ratings for an explicit (button press) or an implicit (letting go of the wheel) method to activate automated mode. Feedback regarding the activation methods, however, showed that a button should be located in the drivers' field of view, and that the movement of the steering wheel made it complicated for the drivers to take back control at the right time. It can be assumed that explicit buttons on the steering wheel may be a good alternative to activate an automated driving system. Otherwise, the retraction of the steering wheel was considered a good mode indication and was thus rated high. The trust results did not show remarkable differences between the groups of participants. Some participants reported difficulties with answering trust items about the interaction with the system or the aid during a decision, because it did not fully apply to the system at hand. The engagement in a non-driving-related activity had a great distracting effect. Drivers that were watching a video or reading a text did not always pay attention to what was presented on the side of the screen, even though the indications were shown close to the infotainment location. As a result of this, drivers of both groups (with or without maneuver indications) asked for more information regarding the automated drive and demanded to be informed (more) about the perception of surroundings and about the actions of the vehicle ahead of time. It seems that the information about the automated driving system should be presented in a prominent and comprehensive way, not casually to make room for other, parallel information.

Summary. A first interaction concept for automated driving was tested in this simulator study. The results give hints as to how such an interaction between driver and system should take place. Positive and negative aspects regarding the activation method and the interface were collected, and consequences for system design were derived. According to the participant's feedback, an activation method needs to be in the field of view, as well as any mode indications and signals from the automated system. These remarks were taken into consideration in the third pre-study.

Pre-study 3

Before implementing a new HMI concept in an actual car, another virtual test was set up to further evaluate the content and location of the HMI. Implemented as a video study, participants experienced different scenarios an automated vehicle could come across, and a dedicated HMI concept for each situation. Before, participants had mentioned that they would like to receive driving related information in the immediate surroundings of the driving situation—thus near the windshield. To investigate if this really would be the preferred solution, also regarding trust in the system, a head-up display (HUD) was compared to a center console display.

Participants. 40 employees of the Electronics Research Laboratory (were the study took place) volunteered to take part in the study. The 32 men and 8 women that participated were not familiar with automated driving before the study. They were on average 29.16 years old ($SD = 6.58$ years).

Method. To investigate the demand of information regarding the driving behavior further, 40 participants were asked to evaluate a new HMI concept. Videos of a highly automated drive were used to illustrate the functionality of the HMI concept. Eight videos of approximately 1 min length each were shown in randomized order. The 2 x 2 study design included a variation of two between-factors, the first one being the level of HMI information. HMI concept A displayed less information, while HMI concept B displayed more detailed information. The other factor examined the position of the HMI, either in the HUD or in the center console (as visualized in Figure 3.5). The additional information in concept B included acceleration and deceleration information, reasons for takeover maneuvers, and other detected vehicles. The videos were shown on a large screen in front of the participants (with a screen diagonal of approximately 55 inch), who were sitting at a desk. The HUD was realized within the videos, whereas the center console screen was represented by a laptop screen next to the large screen. The position of the screens was adapted from the displays in a vehicle.

Questionnaire. Demographic data was collected with a short initial questionnaire (see Appendix A.1.3, Table A.6 and Table A.7). To collect data on how much participants trusted the automated driving system depending on what information is given and where, they were asked to fill out a questionnaire during the study. Each video should be rated regarding participants' trust in the system to handle the situation and their assessment of the system's performance, fol-

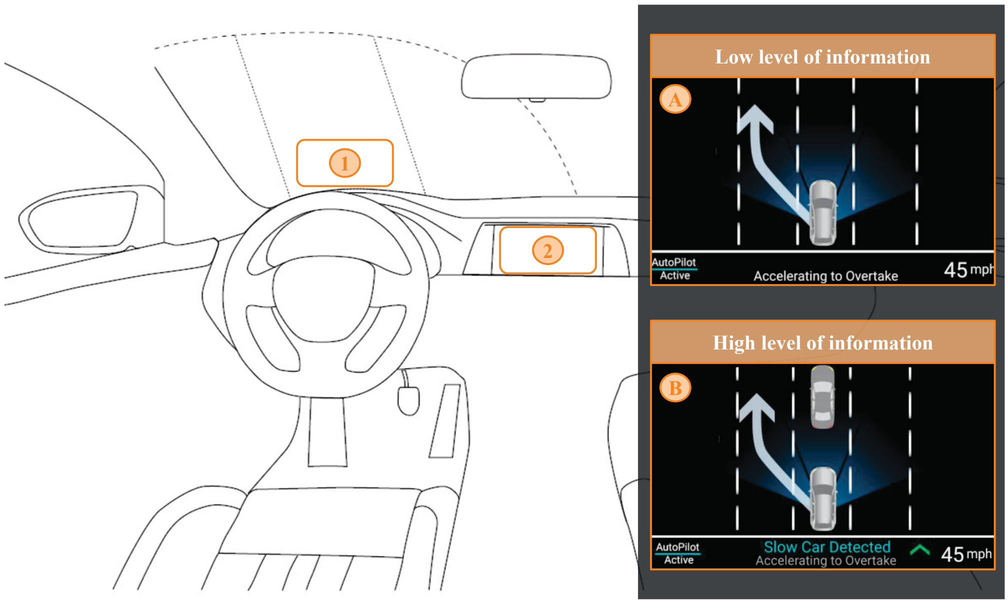


Figure 3.5. HMI concepts of the third pre-study with the factors *level of HMI information* (A – less information, B – more information) and *position of HMI information* (1 – HUD, 2 – Center console display).

lowed by specific questions on the information given in the HMI concepts. For trust assessment, the questionnaire of Jian, Bisantz, and Drury (2000) was used (Appendix A.1.3, Table A.3).

Results and discussion. Trust ratings were related to the system performance ratings in the situations (pearson correlation coefficient $r = .67$, $p < .001$). Furthermore, the performance of the system was rated higher when a HUD was used to display the information for the driver ($M = 12.28$, $SD = 1.75$) compared to the use of the center console screen ($M = 11.23$, $SD = 1.36$). This HMI position seemed to facilitate a comparison of the system's behavior with reality. While this is a helpful insight for further HMI design, a HUD might be more difficult to realize in practice, as the additional display is associated with costs and more technical expense. No striking difference was found in trust ratings depending on the information shown about the automated driving system. This could be due to the simulated environment, which does not create an uncertain situation for participants that would make additional information necessary. Trust, however, is defined as an attitude that is of importance in situations characterized by uncertainty (Lee & See, 2004, see Section 2.2.1). Also, the difference in the HMI concepts was minor (see Figure 3.5), and the exposure time may have been too short to recognize all details. Some participants furthermore reported difficulties rating their system trust—partly because trust items were difficult to interpret with regards to the automated driving system, and partly because the ratings were based on a video sequence. This feedback demonstrates that trust is a construct difficult to assess in a valid way using a simulated environment.

Summary. In the third pre-study, results hinted at the relevance of the location of HMI content for an assessment of the system's performance. The closer the information to the actual driving situation, the better an assessment seems to be possible. The pre-study furthermore showed that for trust ratings to be valid, a realistic setting is of high importance. Besides, questionnaires for trust in automated systems need to be adjusted to match the automated driving system.

Conclusion

Results of these first exploratory user studies highlight research areas that are important to address through further studies. They are, most importantly, strengthening the assumption that detailed system- and driving-related information is needed on the part of the driver. Besides, the qualitative results give hints as to where to display the information (close to the driving scene), how to design the activation of the system (explicit and in the driver's field of view), and what level of distraction of the drivers to expect. Also, it seems necessary to put some effort in the design of a real-driving study to be able to measure trust in automated driving systems in a valid way and to study factors that influence human-automation trust in a real-world environment.

3.3.2 Deduction of interaction concept

To be able to answer the aforementioned research questions and gain insights into the development of trust in automated driving, a research platform was needed. Firstly, as one objective of this work was the investigation of trust under real traffic conditions, a highly automated test vehicle was required. To this end, a highly automated vehicle of Volkswagen Group Research was used for the investigation (Bendewald, Glaser, Petermann-Stock, & Stephan, 2015). Secondly, a dedicated HMI concept for automated driving was utilized to find out how system design can influence the evolution of trust. For this purpose, a special HMI concept was created on the basis of the findings of the initial studies, with particular attention payed to the increase of system transparency. Both premises are described in more detail in this section.

Test vehicle

For the studies under real traffic conditions, a concept vehicle was used, with automated driving functionality as well as a special HMI concept integrated in it. It represents a prototype with SAE level 3 (see Section 2.1.1), where the fallback level is the co-driver.

An automatic driving functionality for highway driving is implemented in the test vehicle, making it possible to use an actual intelligent vehicle for the tests. The hardware and software of the vehicle enable it to gather information about its surroundings via special sensors. It is furthermore able to interpret this information to create a map of the environment and locate itself in it. This way, the automated test vehicle is capable of driving on a highway, at a speed range from 0 to 130 km/h (approx. 80 mph). The system is able to keep the lane and can control the speed as well as the distance to other vehicles. It is furthermore capable of performing lane change maneuvers on its own if necessary. The technical specifications of this functionality will not be covered here, as it is a topic too complex to be described in detail in this work. For an overview, please refer to the official web pages (Audi AG, 2012, 2016).

While the technical side of the automated driving functionality is not in the focus of this work, a closer look will be taken at the display and interaction concept used in the car. A detailed outline of the cockpit of the vehicle can be found in Figure 3.6. The instrument cluster (1) as

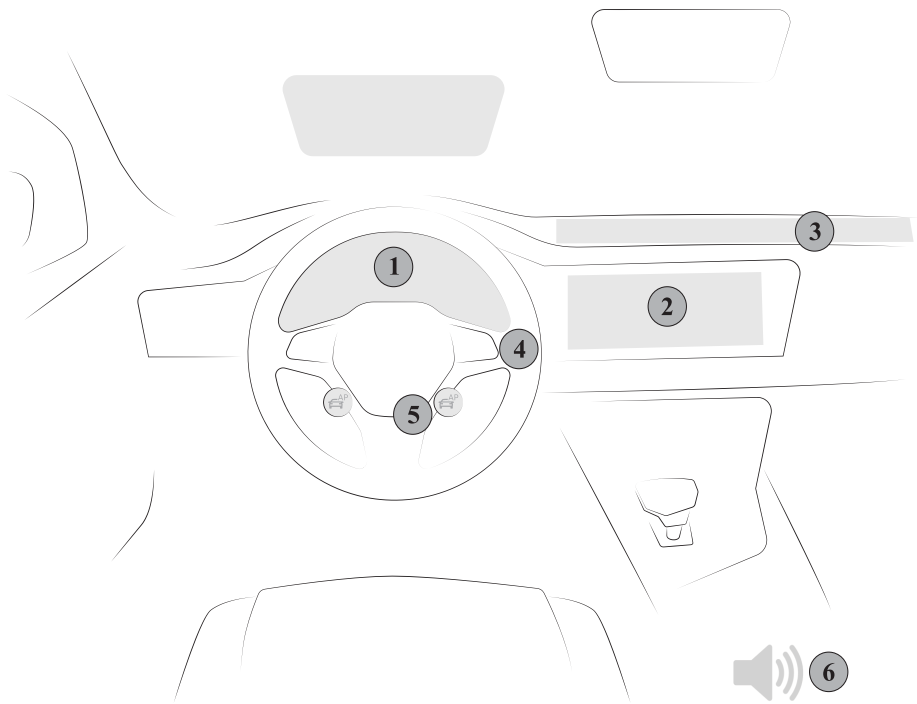


Figure 3.6. Interaction concept of the automated vehicle used in the real driving studies (1–instrument cluster; 2–center console displays; 3–peripheral and ambient lights; 4–flexible steering wheel; 5–two-button concept for activation; 6–speech and sound).

well as the center console displays (2) are the central displays for the driver. They are able to convey basic information about the drive (e.g., tachometer and speed), instructions on how to activate or deactivate the system, and the status of the automated driving system. Whenever the dedicated icon is not gray but white, availability of the automated driving system is indicated, and the elements of the HMI concept change. The instrument cluster shows an icon explaining how to activate the function with two dedicated buttons (5) on the steering wheel, while the LED bar (3) beneath the windshield shows a small turquoise color band indicating that automated mode is available. Certain elements, such as the LED bar and the center console displays, are specifically designed to convey information to a driver who is not in charge of driving anymore and whose attention is diverted away from the driving task. The current state of the automated driving system is always visible to the driver in the LED bar and in the displays. Once the system is activated, the LED bar turns completely turquoise, and the instrument cluster as well as the displays in the center console show an icon in the same color to indicate the new system mode. In addition, the steering wheel (4) retracts to give the driver more space during automated driving. While the automated driving system is active, the steering wheel turns according to the current wheel position whenever a maneuver is performed. A schematic representation of the takeover process can be seen in Figure 3.7. The HMI informs the driver at an early stage (two minutes in advance) of an upcoming takeover request. A speech alert is given to provide the driver with the possibility to prepare for the driving task again and avoid a safe stop maneuver. 15 seconds prior to the manual driving section, all elements of the HMI concept turn orange and escalate to red. A speech command and a distinctive sound (6) indicate the need to take over control.

The user always has the possibility to overrule the system to take back control. To this end, he can use the brake, the steering wheel, or the two buttons on the steering wheel. Each option leads to an immediate handover of complete control back to the driver. During the user studies, a co-driver with an additional pair of pedals always had the possibility to overrule the system in case of an emergency. Apart from these elements designed especially for automated driving, the vehicle's interior was not changed in any way. To make it clear for other road users during the studies that this car is a research platform, it has a distinctive outside appearance, as can be seen in Figure 3.8.

HMI concept

A special feature of the research vehicle is its interface design. It is used to make the system's functionality understandable to the driver. A special HMI concept was designed to increase the automated driving system's transparency and give information to the driver while driving in

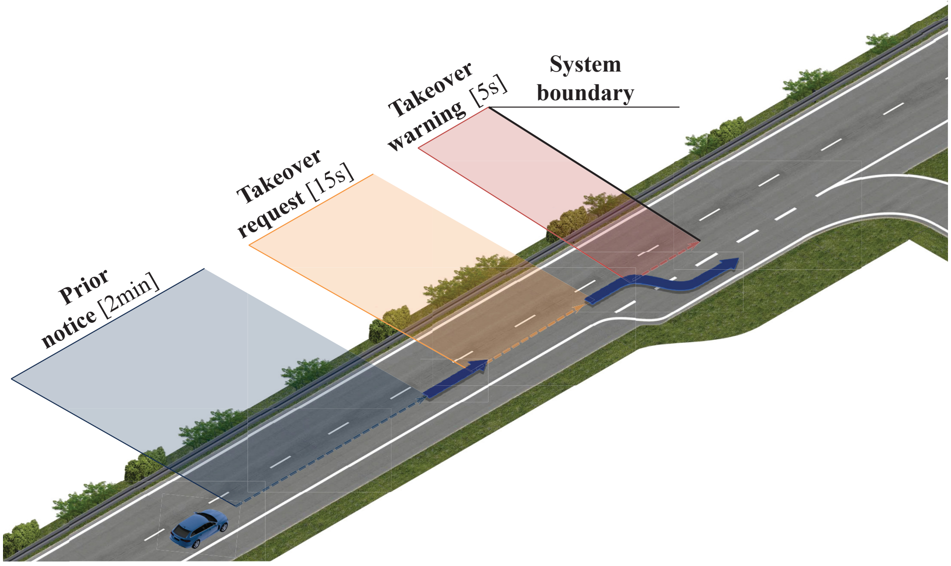


Figure 3.7. Schematic representation of a takeover initiated by the automated vehicle.

automated mode. According to the definition of system transparency by Ososky et al. (2014) (see Section 2.2.3), an HMI was created that made actions of the automated system more transparent to the driver.

As has been shown by Lee and Moray (1992) and Rouse et al. (1992) (see Section 2.3.1), to enhance the driver's mental model of the system, information can be provided on different levels and in different detail. An important mission of human factors research is to find out how information needs to be provided and which information is of particular importance in a certain moment of interaction. When it comes to the level of detail rather than the abstraction of the information (global information about the whole system or concrete information about certain states, etc.; Lee & See, 2004), the appropriate amount of information and the adequate detail level need to be investigated. The amount of information provided has to be balanced carefully: too much information cannot be assessed easily in the short time frame for decision making, but too little information may reduce the usefulness of the indications. As for other automated systems, also for the context of a HAD system it can be assumed that descriptive, explanatory, and predictive indications need to be provided to enable the development of a correct mental model and trust in the system. To understand the system's functionality, its general capabilities and its scope of performance need to be conveyed through *descriptive* information. Furthermore, displaying information about raw data can help to *explain* the system's behavior better. Finally, to enhance the driver's understanding of the human-machine system, he needs to be provided with *predictive* information about the future situation and the actions the vehicle is going to

Figure 3.8. Outside appearance of the concept vehicle used for the real driving studies.

execute. This is especially important because once the driver trusts the automated driving system enough to engage in a non-driving-related activity, he will at least temporarily lose his situation awareness and will be out of the loop of controlling the vehicle. This is deliberate during the automated driving sections to relieve the driver but needs to be overcome quickly if the system is not capable of a certain situation and needs to hand back the driving task. Mode and situational awareness therefore do not need to be maintained but rather reestablished quickly. Examples of such information can thus be the following:

- *Descriptive information* – general capabilities and current mode, e.g.:
    - System state (active / passive / off)
    - Indications to engage / disengage the system (system availability, takeover indication)
- *Explaining information* – sensor information and raw data, e.g.:
    - Localization on the road
    - Detection of surroundings
- *Predictive information* – planned behavior and system boundaries, e.g.:
    - Maneuvers (lane change, taking over), reactions to traffic lights / other cars (braking, stopping, starting)
    - Reason for maneuvers (e.g., overtaking maneuver, obligation to drive on the right) or takeover indications (e.g., missing lane marking, weather conditions)
    - System boundaries (e.g., construction areas, highway junction).

When considering all levels of detail, the feedback out of the system can provide information on the three levels of information described by Rouse et al. (1992) as descriptive, explanatory, and predictive information. For the concept of the transparency display, the aforementioned design recommendations (presented in Section 2.3.2, see Table 2.6) for creating a trustworthy system were taken into account. Also, recommendations of Beggiato et al. (2015) were included, who report that requested information is primarily focused on the status, transparency, and comprehensibility of system actions. Results of their expert focus group highlight the importance of information on system status, remaining time in automated driving mode, fallback level, reasons for a takeover request, and a preview for maneuvers.

To *simplify the interface*, one dedicated display is showing all relevant information regarding the automated driving system. It is displayed in front of the driver in the instrument cluster. The basic version of the HMI concept is designed to only provide status indications of the vehicle (see Figure 3.9). It is used to provide the absolutely essential information to the control groups of the user studies by displaying if the automated driving system is available or active and a prediction for how long it will be available before the driver has to take back control.

Figure 3.9. Interface concept 1 with a status indication.

The second HMI concept furthermore includes sensor and processing information and information about the surrounding environment detected by the vehicle (see Figure 3.10). This information can serve to implement the design guideline to *provide the user with ongoing feedback*. As the detection of the system is displayed, it also *provides access to raw data* as recommended. The user can observe the data collected by the sensors and can thus check the quality of the vehicles perception. Furthermore, these indications of perception and processing of the automated vehicle can help with the conveyance of system limits and boundaries. For example, the overview of detected objects can also help to *provide means to indicate unreliable data*. The driver can observe the reliability of the detection rate, and can decide whether it is still safe to use the automation. It can also hint at a possible *automation boundary*, if reliability is low during a time period.

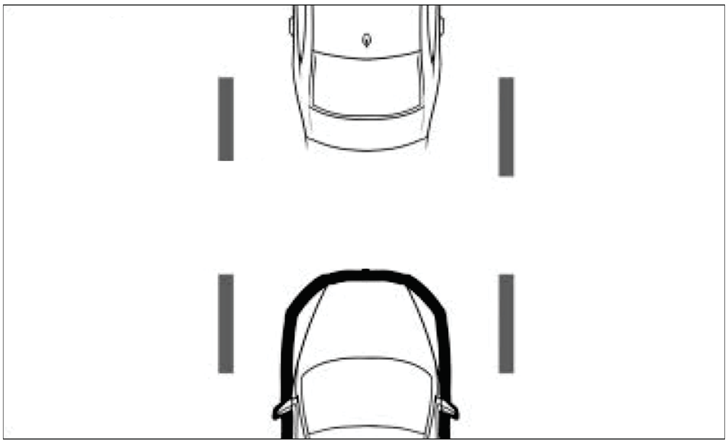


Figure 3.10. Interface concept 2 with surrounding environment (in addition to the status information).

In HMI concept 3 also information regarding the interpretation of the vehicle is presented to the driver. It contains status information, every object the car detected, and all planned behavior of the system in a unified display. The comprehensive interface design can be found in Figure 3.11. Not only are the surroundings displayed, but also additional information on the vehicle's behavior like planned maneuvers of the automated driving system are shown. Every maneuver is announced two seconds in advance and an arrow is displayed in front of the outlined ego vehicle in the direction strived for. These details can help the driver to anticipate the behavior of the vehicle. Also, in order to *reveal the rules and algorithms used by the automation*, the planned maneuvers of the vehicle are supported by an explanation of this behavior. For example, the vehicle will indicate a lane change to the left, ideally supplemented by an explanation for the maneuver (e.g., "overtaking maneuver").

Other principles of design can be implemented in the vehicle's behavior rather than in a single display. Thus, the *purpose of the automation is made clear* through the general interaction principles of the test vehicle. For example, the steering wheel is retracting whenever control is handed over to the system to indicate the shift of responsibility. This also applies to the recommendation to *implement good computer etiquette*. The interaction principles of the automated vehicle include a pleasant female voice announcing upcoming takeovers, giving a positive impression of the system's communication style. Lastly, it was recommended to *train the operator* of the automated driving system. In other domains like aviation, it is crucial and thus

Figure 3.11. Interface concept 3 indicating a lane change to the left (in addition to the status information).

self-evident to train the operator, but in the case of automated driving, standards need to be developed. Drivers need to learn about the system's capabilities as well as about its limits to be able to evaluate the system's performance and identify situations of low reliability. During the user studies, the experience of the drivers was varied to understand the importance of training in this domain.

These HMI concepts were translated into designed versions for the user studies. The different stages were used to determine the influence of system transparency on trust in automated driving by varying the information given in the display. They were implemented in the aforementioned test vehicle as well as in a mock-up for testing in simulated environment. The next chapter describes the approach of three different experimental procedures that were designed to answer the proposed research questions.

4 Studies

Three studies were carefully designed to answer the research questions of this work. On the theoretical side, the objective was to identify factors influencing trust in this domain. On the practical side, indications of the HMI needed to be found that are useful to create or increase trust in an automated vehicle. The user studies were conducted using different HMI concepts, either in a simulated environment or in a real driving study. This variation of system indications was used to get insight into what information is most necessary for the driver to develop trust in the automated driving system. Figure 4.1 gives an overview over the study objectives.

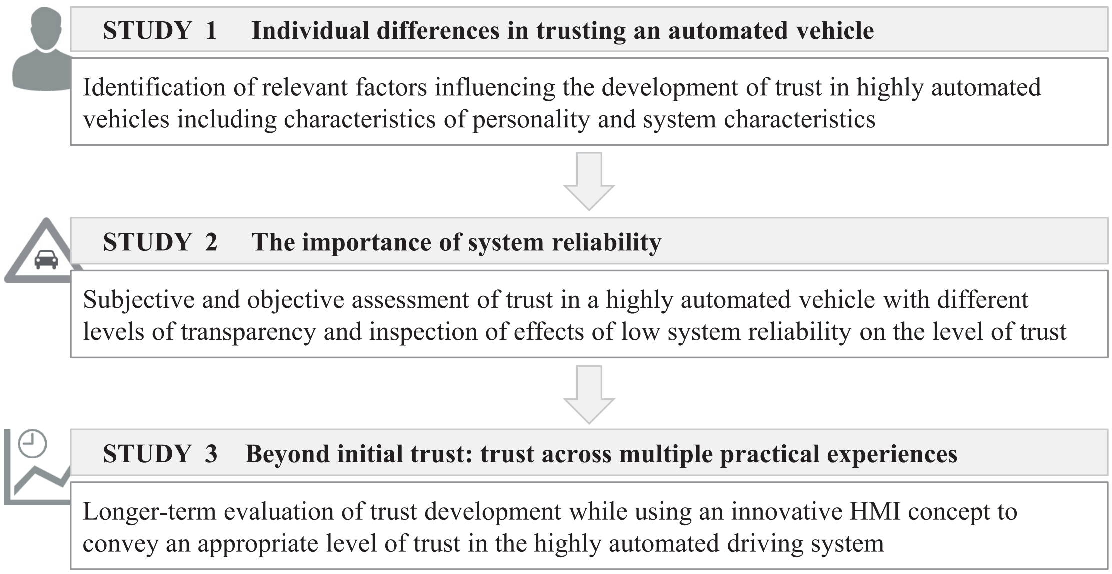


Figure 4.1. Overview over the main study objectives of this work.

This work puts its focus on ways to create and analyze trust in HAD. It concentrates on the reconciliation between the two worlds of the rising capabilities of automation and the users that are still needed as a fallback solution in case of system limits. Nowadays, computers can do more and more tasks on their own—but the users' demand to understand and to be able to use the system also cannot be overlooked.

In this chapter, the three main studies and their results are described in detail. Section 4.1 briefly introduces the methods that were used in all of the three user studies. It gives an overview over the variables of major importance and the means of measurement that were applied. Section 4.2 describes the first user study with the objective to identify individual differences of trusting an automated vehicle. In Section 4.3, the level of trust in automated driving in the event of low system reliability is tested in another user study. Finally, in the last study the development of trust in an automated vehicle across multiple practical experiences is in focus, as described in Section 4.4. Each study section consists of sub-sections on hypotheses, study design, results, and implications.

4.1 Methods

So far, only few studies have used HAD systems rather than driver assistance systems (for instance ACC) or decision aid systems to investigate research questions related to trust in automated systems, and even less can report data from an actual vehicle instead of a simulation. The objective of the first user study was to find out the central and most relevant factors influencing trust development in HAD in a real-world environment. This approach can help to confirm study results of other domains as well as from simulator studies, and can enhance knowledge in this area. In the second user study it was evaluated how trust can be maintained even in the event of low system reliability. To determine the impact of system boundaries on trust development and consistency, the controlled environment of a simulator was used. A longer-term driving study was used to expand knowledge about the evolvement of trust even further. In this third user study, development of trust was investigated across several experiences with the system to find out how a specific HMI concept may help to influence system trust on the long term.

4.1.1 Independent variables

In the studies, different factors were varied to learn how they impact trust in an automated driving system. This paragraph gives a short overview over the main factors taken into account. Mainly, system related factors were altered within the scope of the user studies. Additional factors were added depending on the study hypotheses.

Transparency of the system. In the user studies, it was distinguished between different levels of information provided to the driver in order to find out how detailed the information has to be. To convey these levels of information through the system, HMI concepts with according content needed to be created. The design attempted to take the recommendations of Section 2.3.2 into account. The resulting HMI designs led to different levels of system transparency. They were implemented in different ways for each of the studies, depending on the technical preconditions and the focus of the particular study. The designs were all derived from the interface concepts presented in Section 3.3.2.

System reliability. As this work concentrates on the levels of conditional and highly automated driving (SAE International, 2014), no errors of the automated system are expected. System boundaries exist, but are announced in advance with an adequate time reserve for the driver to react. Still, situations may occur that feel odd for the driver or where he might feel the need to intervene (although he does not need to). An example could be short swerves of the steering wheel or reinterpretations of the driving situation that lead to the cancellation of a maneuver. These unanticipated system reactions are further on referred to as *low reliability situations*. The user studies investigated the influence of such events on trust in the system. In the second study, this was done in a standardized environment, where other possible influences can be held constant. It has been proven before that system reliability affects trust. In the studies at hand, the interaction effect of reliability with system transparency was of interest.

Experience with the system. System experience is a factor difficult to investigate in the research area of automated driving, as the technology is so novel. Not many people have had the possibility to drive in a highly automated vehicle so far, thus long-term studies on the topic of trust in such a system are also rare. One longer-term simulator study conducted by Large, Burnett, Morris, Muthumani, and Matthias (2018) focused on activities during automated driving, with trust being part of the evaluation. The authors found high levels of trust in the system, but they also note that the results may be confounded by the low-risk perception in the driving simulator. The user studies at hand tried to assess the development of trust depending on system experience using different strategies. A detailed familiarization with the system was used as well as an observation over several practical experiences.

4.1.2 Dependent variables

To analyze consequences of the variation of the independent variables, subjective as well as objective variables are of interest. The following section presents an overview over the dependent variables. The methodology of measurement can be found in Section 4.1.4.

Subjective data. Trust ratings, the assessment of system performance, as well as assessments of the usefulness of the indications provided by the automated driving system were part of the subjective measures in each study. The detailed description of the data collection can be found in the respective chapters of the studies (Sections 4.2.2, 4.3.2, 4.4.2).

Objective data. Behavioral measures were taken into account as well. Objective data were obtained by measuring the handover and takeover times (reaction times to handover and takeover requests) to assess the driver's readiness to use the system and his ability to resume the driving task. This was also used to get an estimation of the level of situation awareness. In addition, the gaze behavior of the driver was used to give insight into trust as an objective measure. Numerous control-glances to the surrounding traffic situation, for example, could suggest that the driver feels the need to supervise the vehicle, thus showing low trust to a certain degree. The use of non-driving-related activities while being driven by the automated driving system was also used as an indication for trust. Physiological data was collected to verify if trust can also be inferred from certain physiological reactions.

4.1.3 Mediating variables

While the specification of the aforementioned independent factors can be actively varied in an experiment, other factors relevant for trust development might not be modifiable. Those covariates were collected because they can potentially mediate effects occurring as study results. They were included in the analyses in order to identify the underlying processes of trust formation.

Demographic data. Demographic data usually includes information about age and gender of the participant. In the context of automated driving, information about driving experience or experience with driver assistance systems can also be of interest and was collected with a questionnaire (see Appendix A.1.3, Table A.6). In this context, participants were also asked to describe their driving style and rate themselves as drivers compared to others (adapted from De Craen, Twisk, Hagenzieker, Elffers, & Brookhuis, 2011).

Personality traits and attitudes. Each study included measurements of personality characteristics and attitudes. Factors in focus were attitudes like the personal attitude towards technology, but also personality traits like extraversion, desire for control, the tendency to take risks, and self-efficacy (see Section 2.2.3). The means of measure for each of these variables are detailed in the following Section 4.1.4 (see also Appendix A.1.3).

States. To also capture the current state of the participants, their subjective level of stress was measured during each test drive (see Appendix A.1.3, Table A.8). Stress was the personal condition that had been most important in the pretests.

4.1.4 Methodology of measurement

To further enhance the understanding of the process of trust development in automation, trust needs to be made measurable. Many years of research on trust assessment provided different approaches, using subjective as well as objective methods for an evaluation. The methods used in the user studies to collect the relevant variables are presented here.

Subjective measurement of trust

Trust, considered a mental state, is mostly measured through *subjective* appraisal (e.g., Llinas et al., 1998; Muir, 1989). The concept is generally expected to be assessable internally, as it is based on rational and emotional processes (Wang, 2010). When using appropriate scales, subjective ratings can help to obtain reliable and repeatable data (Moray & Inagaki, 1999). Subjective ratings were found to be sensitive enough to discriminate between changes in the properties of a system (Muir & Moray, 1996). To assess trust in an automated system, often the human-human relationship is taken as a point of reference (e.g., Rempel et al., 1985). The Interpersonal Relationship Scale developed by Rotter (1967) is still a common method to measure trust also in the context of human-automation interaction. However, as the field of research on trust in automation is growing, meanwhile several scales were developed exclusively for this new context. Several attempts have been made to measure trust in automated systems with the help of a questionnaire. A short overview shall give an idea of the different approaches, presenting a selection of questionnaires assessing trust in automated systems.

A well-known trust questionnaire was developed by Jian et al. (2000) and validated by Safar and Turner (2005) and Spain, Bustamante, and Bliss (2008). The authors created a checklist for trust between people and automation during an interaction, assuming that trust is a continuum ranging from trust to distrust in a two-dimensional structure. The twelve items on the sub-scales trust (seven items) and distrust (five items) were rated on a seven-point scale (see Appendix A.1.3, Table A.3). The System Trust Scale is assumed to assess a general attitude toward automation, but not trust in a specific system. Other questionnaires, like the Human-Computer Trust instrument developed by Madsen and Gregor (2000, see Appendix A.1.3, Table A.4), or the trust scale by Wiczorek (2011) include specific items about the interaction between system and human. They are thus rather aiming at measuring trust in assistance systems, lower levels of automation, or advisory systems. Muir (1989) was one of the first to create a scale exclusively for assessing trust in an automated system helping with a laboratory task. In her dissertation, Muir developed a scale for automation trust to test the models developed by Barber (1983) and Rempel et al. (1985). The items were assessed on a line with the poles ‘not at all’ and ‘extremely’ to measure the perceived competence, predictability, dependability, and responsibility of the system. In addition, faith in the future performance of the system, trust in the reliability and in the system's display as well as overall trust in the system was collected. All these measures correspond directly to the dimensions of trust proposed by Muir (1989) (see also Muir & Moray, 1996). A shortened version of this questionnaire was applied in later research. In their experiments, Lee and Moray (1992) measured operators' level of trust in a supervisory control system with a questionnaire modeled on the items introduced by Muir (1989). On a ten-point scale, participants were asked to rate predictability, dependability, faith, and trust in the system, corresponding to the trust dimensions proposed by Muir (1989) (see also Desai, 2012). The questionnaire and the self-translated items can be found in Appendix A.1.3, Table A.5.

Most of the questionnaires regarding trust in automation either address a general attitude toward automated systems or focus on assistant systems like decision aid systems supporting the human in a task. To use a scale for trust addressing a specific system as well as a high level of automation with a complete shift of tasks to the system, the shortened version of the questionnaire on trust in automation (see Desai, 2012; Lee & Moray, 1992) was applied in the work at hand (originally developed by Muir, 1989). It consists of four questions, developed to collect a one-dimensional rating for trust in an automated system:

- To what extent can the system's behavior be predicted from moment to moment? (Predictability)
- To what extent can you count on the system to do its job? (Dependability)
- What degree of faith do you have that the system will be able to cope with all situations in the future? (Faith)
- Overall how much do you trust the system? (Overall trust)

These items were used for the user studies due to their simplicity and applicability in an elaborate study setting in the context of automated driving. They were assessed with a 15-point rating scale based on Heller (1985) (see Appendix A.1.2, Table A.2), ranging from 1 = 'very low' to 15 = 'very high'.

Behavioral measurement of trust

Another possibility to assess the level of trust of a person in a system is the *objective* measurement of trust-related constructs as an operationalization of the actual construct. In contrary to subjective measures, objective measures are more related to reliance than to trust, to be accurate. These constructs are understood as the behavioral outcome of trust. They are related closely to trust and are thus of high interest as well.

Use of the system. One possible operationalization may be the use of a system. Assuming that a system will only be used if trusted, the decision to use the system as well as the frequency of use might be good indicators for trust in the automation. Studies of Lee and Moray (1992, 1994) as well as Muir and Moray (1996) used the amount of time spent using an automation as an indicating variable and were able to prove a high positive correlation between trust in and use of the automation.

Takeover times. Reaction or takeover times can be a useful measure of internal processes, because the reaction examined is often an involuntary response to a certain stimuli and is thus more difficult to tamper. A study on complacency by Knapp and Vardaman (1991) provides results regarding the reaction time to a warning in a controller maintenance task. Participants waited for a reaction of the automated system before acting themselves, which is interpreted by the authors as a high level of complacency. A recent study by Pradhan, Ranjan, and Samal (2015) confirmed this finding in a study on reaction time depending on reliability of the automated aid. Results showed a significant longer reaction time when automation was highly reliable compared to a 25% reliability condition, indicating that reaction time can be used to predict trust in an automated task. Because of the driver being out of the loop of controlling the vehicle, takeover times when automation requires to be replaced by manual steering are one main focus of research (e.g., Gold, Damböck, Bengler, & Lorenz, 2013; Gold, Damböck, Lorenz, & Bengler, 2013) and are reported to range up to 8.8 seconds (Petermann-Stock, Hackenberg, Muhr, & Mergl, 2013). As the driver is allowed to withdraw attention from the driving task during highly and conditionally automated driving, it is advised that the automated driving system is equipped with a facility that maintains functionality for at least ten seconds as a technical fallback solution (Petermann-Stock et al., 2013).

Monitoring behavior. Not only can the actual use of the system be helpful to infer the current level of trust. When focusing on trust in automated driving, it might be possible to get an impression of how much the person trusts the vehicle by looking at their monitoring behavior (Popken, 2009) and their voluntary distraction from the driving scene. As a consequence of the definition of HAD, it is fully up to the driver to what extend he monitors the vehicle while driving in automated mode. It can be assumed that the amount of monitoring directly depends on the level of trust in the system. Direct evidence for this assumption has recently been provided by a study in a driving simulator (Hergeth, Lorenz, Vilimek, & Krems, 2016). Based on eye-tracking analyses, the authors report a medium to high negative correlation between self-reported trust in the automated vehicle and the amount of monitoring the drive, the latter operationally defined by the monitoring ratio, i.e. percent of gazes directed to mirrors, instrument cluster, or windshield. Other research has provided insights into the effectiveness of assessing the level of trust via gaze data before. Muir (1994) predicts that automation that is highly trusted will be monitored less because uncertainty is low and a close observation is expected to be unnecessary. Results of Muir and Moray (1996) prove the inverse relationship between trust and monitoring behavior. Hence, most monitoring is expected to happen with medium levels of trust or when the operator is not yet familiar with the system and uncertainty is thus still high (Adams et al., 2003). Sheridan and Hennessy (1984) assumed that operators with low trust in a system will spend more time monitoring the automation when they (have to) use it. This assumption was confirmed by experiments conducted by Muir and Moray (1996), who found an inverse relationship between trust and monitoring. Thus, trust as a psychological state was proven to be a factor causing diminished monitoring behavior. A driver trusting an automated vehicle to manage all situations will not check on it too often by looking in the mirrors or following up on the surroundings—he will rather keep his eyes on a non-driving-related activity or let them wander around without a definite destination. The fact that trust and monitoring are so closely linked can help to find an objective assessment of trust. To give insight into trust as a behavioral indicator, analysis of gaze data can show control-glances to the surrounding traffic situation or the mirrors. This monitoring or information-sampling behavior can be interpreted as supervision that shows distrust in the system to a certain degree. Several researchers looked into the topic of attention allocation and eye glance behavior as an objective measure of automation reliance (e.g., Hergeth et al., 2015; Parasuraman & Manzey, 2010; Sheridan & Parasuraman, 2005). They assume a close relation between attention and gaze movement (Zeeb & Schaub, 2014). Moray (2000) in this context define a human monitoring a system more frequently than necessary (or optimal) to be *sceptical*. A human monitoring a system less frequently than optimal is called *complacent*. Metzger and Parasuraman (2001) arrive at the same result when trying to measure trust via gaze behavior (here the amount of checking the display during an operator control task). They found a decrease in attention allocation compared to the manual condition caused by overreliance on the automated aid.

Summing up the research results, it can be expected that people will spend more time checking on the automation when they have no sense of security that the automated system will act as they want it to (Sheridan & Hennessy, 1984). It is assumed that the analysis of gaze behavior can give insight into trust in automation, indicating distrust when people supervise the automated system's behavior (in this case the automated driving system and the driving scene). People not trusting the system are expected to show a scanning behavior similar to manual driving (Gold et al., 2015), while trusting people are assumed to divert their attention away from the driving task. This assumption is also related to the use of side tasks when trusting the system.

As a registration device for allocation of attention, the head-mounted Dikablis eye-tracking system (by company Ergoneers GmbH, 2015) was used for the research at hand. It allows data collection and processing of data. Specialized cameras are mounted to an eye-tracking frame the participant needs to wear similar to glasses. The mount allows the perception of the environment with only minor impairments. The left pupil is tracked with a camera facing the eye of the participant. Another camera is facing forward and is capturing the field of head direction to record the surroundings. The associated software D-Lab then links gaze data to specific areas of interest (AOIs). Figure 4.2 shows the AOIs used in the user studies: glances to the surrounding traffic situation (street), the side and the rear view mirrors, the instrument cluster, and the center display. An additional area for non-driving-related activities was defined below the steering wheel (not marked in the figure), where drivers tended to use their smartphone. To calculate the percentage of gazes on the different AOIs, the standard output of the eye tracking software D-Lab Ergoneers GmbH (2015) was used.

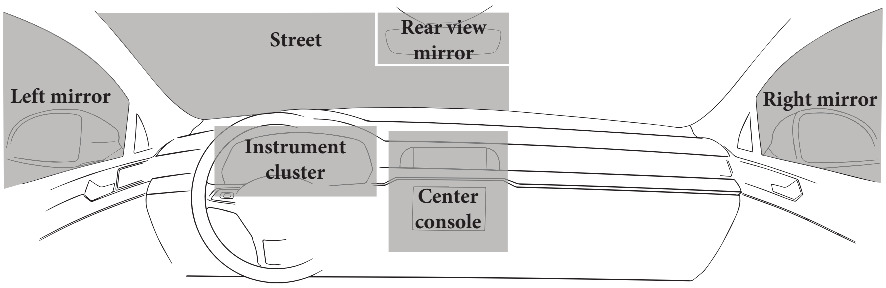


Figure 4.2. Areas of interest within the cockpit used for the collection of gaze data.

Use of non-driving-related activities. An indication of trust in an automated vehicle can certainly be the time spent with other activities or devices while being driven. Buld, Tietze, and Krüger (2005) describe a withdrawal of attention and a tendency to deal with non-driving-related activities as an effect of replacing assistant systems. Related to gaze data, the use of non-driving-related activities can thus show that the person is not involved in the driving task anymore, having handed the control of the vehicle entirely to the system and not supervising it anymore. In their experiments, Helldin et al. (2013) found that people engaged more in non-driving-related activities while still being able to take over the driving task best (and thus react fastest to takeover requests) when trust is calibrated in an appropriate way.

Mediating variables

Control variables were collected in the studies to find out which factors influence trust in the context of automated driving (for a description of these characteristics of the human, see Section 2.2.3). The personality and attitude questionnaires were assessed with a 5-point Likert-type rating scale (see Appendix A.1.2, Table A.1) ranging from 1 = 'strongly disagree' to 5 = 'strongly agree'.

Demographics. Demographic measures included gender, age, years since driving license, regularity of driving, and experience with advanced driving assistance system. A short questionnaire was developed to capture these elements (see Appendix A.1.2, Table A.6 and Table A.7).

Driving style. Also in the initial questionnaire, it was collected how people perceived themselves as drivers, comparing their own driving capabilities to others. The questions were shortened and adapted from De Craen et al. (2011). To get an idea of how people liked driving, their preference of being driver or passenger was collected with another question designed by the author. The questions can be found in Appendix A.1.2, Table A.7.

Extraversion. Individual differences in diverse personality factors were assessed. For the psychological factor extraversion, the short form for the International Personality Item Pool (IPIP) representation of the revised NEO Personality Inventory (NEO PI-R) was used as a means of measure (Johnson, 2006, see Appendix A.1.3, Table A.9)). Extraversion in this questionnaire consists of the subscales warmth, gregariousness, and positive emotions. The twelve items need to be rated regarding the degree of consent with the sentences on a 5-point rating scale (see Appendix A.1.2, Table A.1).

Desire for control. Burger and Cooper (1979) developed a scale to measure a desire for control. The 20 items consisted of sentences regarding control in various contexts, ranging from political participation to handling different situations in life (see Appendix A.1.3, Table A.10). The questions were later found to load on the three factors *desire for leadership and independence* (control others), *desire for not having to take decisions* (relinquish control), and *desire for determining own life* (control self) (Gebhardt & Brosschot, 2002). Originally used with a seven-point scale to answer the items, here the 5-point scale was applied to keep the same scale for the complete initial questionnaire (see Appendix A.1.2, Table A.1).

Tendency to take risks. To assess participant's tendency to take risks, a four-item questionnaire was used that was initially employed by Lee and Moray (1992) as well as Desai (2012) (see Appendix A.1.3, Table A.11). This questionnaire originally used a six-point scale, but was adapted to fit to the structure of the other questionnaires and to not confuse participants with different scales. Thus, the 5-point rating scale was again utilized (see Appendix A.1.2, Table A.1).

Self-efficacy. A general self-effectiveness or self-confidence was assessed as well. As a means of measure, the self-efficacy scale by Schwarzer and Jerusalem (1995) was utilized (see

Appendix A.1.3, Table A.12). In this scale ten items regarding coping strategies with problems and behavior in difficult situation need to be answered. The four-point scale was converted into a 5-point scale again (see Appendix A.1.2, Table A.1).

Acceptance of technology. Acceptance of technology was collected with a questionnaire introduced by Karrer et al. (2009). In this questionnaire, declarative sentences on interaction with technical devices shall be answered, again on a 5-point scale (see Appendix A.1.2, Table A.1). The items can be structured regarding the subscales enthusiasm, competence, positive attitude, and negative attitude (see Appendix A.1.3, Table A.13).

Additional variables

Other subjective data was gathered with the help of single items, in order to find out more about the user's experiences with the automated driving system. As with the trust items, the additional items were rated on a 15-point scale ranging from 1 = 'very low' to 15 = 'very high' (based on the idea of Heller, 1985, "categorial subdivision procedure") (see Appendix A.1.2, Table A.2). The specific items can be found in Appendix A.1.3, Table A.8.

Nervousness. Participants rated their nervousness during the automated test drives. This subjective assessment was assessed to understand participant's level of stress in this novel situation.

Perceived performance of the system. During and after interacting with the system, drivers were asked to rate their perceived performance of the system.

Evaluation of HMI. The usefulness of the information given by the interface was retrieved, including a rating of the preferred HMI version drivers would like to use.

This section introduced the methods that were utilized in this research to answer research questions related to trust in an automated vehicle. Measurands utilized in other research were presented that can be feasible for this context. The following sections present the three user studies and their results in detail.

4.2 Study 1: Individual differences in trusting an automated vehicle

The first user study was conducted to examine the impact of different determinants on driver's trust in a HAD system. To shed light on potential factors having an impact on trust in this context, different factors previously shown to influence trust in other domains were taken into account. While factors like personality characteristics and attitudes potentially have an effect on dispositional trust in the system, usage of the system is assumed to affect situational and learned trust (see Section 3.1). Specific factors were added to see if trust could be guided in a controlled way through this interaction.

As a first study of its kind, the driving study presented in this section aimed at investigating the topic of trust in automated driving "in the wild" (Hutchins, 1995). It was conducted in real traffic conditions and involved participants actually driving with an automated test vehicle to explore relevant factors for trust development in intelligent vehicles. A prototype automated vehicle of Volkswagen Group Research (Audi A6 Avant) was used as a research platform for this purpose (Bendewald et al., 2015, see Section 3.3.2).

4.2.1 Hypotheses

The foregoing theoretical considerations resulted in research questions for the first user study. According to literature, individual differences in trusting an automated vehicle can be based on a variety of factors. In this study, a closer look was taken at personality characteristics of the driver in particular.

As has been described before, research on dispositional trust has identified several personality characteristics and attitudes to be relevant for the formation of initial trust in an automated system. This led to the first hypothesis.

> *Hypothesis 1: Dispositional trust in an automated driving vehicle is dependent of the driver's personality characteristics (e.g., extraversion, tendency to take risks, desire to be in control, perceived self-efficacy) and attitudes towards technology in general (e.g., acceptance of technology).*

Apart from demographic factors like gender and age, personality factors and attitudes seem to play an important role in the development of trust, whether it is in other humans or in automation. For the context of HAD, the following assumptions were made resulting from former research:

- *1a) Initial trust of the driver will be positively affected by a high acceptance of technology.*
- *1b) Initial trust will be higher for an extraverted person.*
- *1c) Initial trust of the driver will be positively affected by a high risk-taking behavior.*
- *1d) Initial trust will be lower for drivers with a high desire for control.*
- *1e) Initial trust of the driver will be lower when the person has a high perceived self-efficacy.*

In other domains of automation, the situational context was found to be relevant for reliance on automation Lee and Moray (1992, 1994). In the driving study presented here, people were asked to use the system, and could not allocate task responsibility based on their own decision. However, it was postulated that the situational context also plays a role for trust in HAD. It was assumed that participants will trust a HAD system less during complex driving tasks, because drivers do not know the capabilities of the system yet.

> *Hypothesis 2: Situational trust in an automated driving system will be lower in complex driving situations.*

It is evident that machine characteristics are highly influential factors regarding trust in automation. This was also assumed to be the case in the context of automated driving. On the one hand, gaining experience with an automated system has been proven to increase trust in a machine, and transparency of the system can promote this trust further by reducing uncertainty (Verberne et al., 2012; Ye & Johnson, 1995). On the other hand, trust is expected to change when experiencing system limits or unexpected system behavior (Manzey et al., 2012). When the automated driving system reacts in an unintended way or has to give control back to the driver, this might entail diminishing effects on trust.

> *Hypothesis 3: Learned trust in an automated driving system develops based on the use of the system, the experience of the system's capabilities and boundaries, and its transparency.*

In general, it was assumed that while gaining positive experience with an automated driving system, just like with any other automation, people will develop trust in the system. The more a person is interacting with a certain system, the better will the person's mental model of the system be. A correct mental model, in turn, has been assumed by Kazi et al. (2007), Itoh (2012), as well as Beggiato and Krems (2013) to be the basis for adequate trust in automation (see Section 2.2.3).

> *3a) Trust will be higher with rising positive experience with the automated driving system.*

Going along with the hypothesis that trust development depends on the shaping of the mental model, it was furthermore assumed that this process can be supported. Specifically, the system's transparency is expected to have a major influence on the perception of the system and the trust that is put in the system. Research has been suggesting that providing information about the systems capabilities, boundaries, and intended actions can help making the system more transparent, thus facilitating the development of a correct and comprehensive mental model of the system.

3b) *Trust will be positively affected by providing information about the system's state and behavior to the driver, thus creating a more transparent system.*

In this context, a positive effect was understood as an enhancement of trust to an appropriate level. Overtrust could of course also arise, making it necessary that the information given about the system is truthful and contains hints about system features as well as boundaries.

Trust in automated driving is not only expected to manifest itself in a subjective believe, but also in drivers' reliance behavior. Specifically, high levels of trust are assumed to lead to longer periods of driving without monitoring the drive. Direct behavioral indicators of such attentional shift might be the time spent with other tasks than supervising the drive (e.g., reading, using smartphone), or the number of control gazes to mirrors or the instrument panel to check for proper driving (Hergeth et al., 2016; Zeeb & Schaub, 2014).

Hypothesis 4: Trust in an automated driving system goes along with a shift of attention away from the driving situation as reflected in dealing with non-driving-related activities and/or a reduction of monitoring of relevant aspects of the driving task (e.g., speed, other road users).

People that have a higher level of trust were expected to divert their attention from the driving scene, as they do not feel the need to supervise the system while it is taking control of the vehicle for them. They were supposed to be more likely to show complacent behavior (see Section 2.2.2), relying on the system and not monitoring its actions. Direct behavioral indicators of trust might be the number of control gazes in the mirrors or the instrument panel to check proper functioning of the automation (Zeeb & Schaub, 2014), or the time spent with other tasks than supervising the system (e.g., performing non-driving-related activities like reading or using the smartphone).

4a) *A person trusting the automated driving system more will check up on the system less often.*

4b) *With a higher level of trust, a person will avert more from the driving situation.*

As the driver is still the last fallback level for the HAD system in the event of a system limit, the vehicle may prompt the driver to take back control. Depending on his level of trust, the driver might not pay attention to the traffic situation and may thus be out of the loop of driving. Hence it was suspected that such a driver will need more time to react upon an unexpected takeover request from the vehicle.

4c) *A person with a higher level of trust in the automated driving system will have a higher takeover time.*

These hypotheses were intended to be verified or disproved by the first user study presented here. In order to investigate them, a suiting study design was created, as presented in the next Section 4.2.2.

4.2.2 Study design

A user study under real traffic conditions was designed to test the aforementioned hypotheses. An evaluation of characteristics of the driver as well as the system was conducted to assess how trust in automated vehicles evolves and how this development is affected by certain characteristics.

Participants

Participants of the first user study were recruited internally and were all employees of Volkswagen Aktiengesellschaft in Wolfsburg (Germany). They participated voluntarily and received a gift after the experiment as an incentive for taking part in the study. 8 women and 20 men took part in the driving study. They had an average age of 36.61 years ($SD = 9.37$ years). Their average of 26 232 km of driving per year ($SD = 20527$ km) showed that they were experienced drivers. Furthermore, most of the participants were familiar with driver assistance systems like cruise control, adaptive cruise control, or heading control and used them regularly. However, none of them was familiar with HAD systems or involved in the development of it. Of all participants, only three stated to enjoy driving as a passenger. All others preferred driving themselves and explained this with the fun of driving, but also with being more comfortable when having control over the situation. Most drivers considered themselves as better drivers or equal to average. Only two participants thought of themselves as worse drivers than others.

Test vehicle

For the first main real driving study, the prototype concept vehicle with HAD functionality described in Section 3.3.1 was used as a research platform (see also Bendewald et al., 2015). The vehicle was equipped with the capability to drive highly automated on highway roads at speeds ranging from 0 to 130 km/h (approximately 80 mph). That is, it could control speed and distance to other vehicles in the front, keep the lane and perform maneuvers such as automated lane changes. The interaction procedure that was designed especially for the automated driving functionality was used (see Figure 3.6). The vehicle's interface included a unique color indicating system availability and status in an LED bar below the windshield, and a specific icon in the instrument cluster. Special situations like system availability or an upcoming takeover were emphasized by a distinct tone. Preceding a takeover request, a voice indicated that manual driving was required in two minutes. Two HMI versions displayed in the large center console display were compared in the study and are described in the following.

Study design and independent factors

A 2 (system transparency) x 2 (situation complexity) mixed factorial design was used. The first factor was a between-subjects factor with two levels corresponding to different HMI versions developed to inform the driver about the automated system. The information given in the interface was varied in order to compare different levels of transparency of the automated driving system and to be able to assess the impact of system transparency on trust in an automated driving system. In combination with the factor situation complexity, the study's goal was to clarify whether the HMI elements are gaining importance in certain traffic situations. The second factor was defined as a within-subjects factor and included varying situation complexity. The complete design of the study is visualized in Figure 4.5.

System transparency. Two different versions of an HMI were designed, making the system's behavior more or less visible to the driver and thus implementing a lower or higher transparency of the automated driving system. In the low-transparency condition A (see Figure 4.3, left), the road with lane markings and the ego-vehicle were displayed as well as other surrounding cars detected by the system. All processing and interpreting done by the system was not made visible to the driver. In the more comprehensive version B (see Figure 4.3, right), additional information was given to the driver. The information made the automated system's behavior more transparent by displaying intended actions (i.e., lane changes or overtaking maneuvers) with an arrow. Possible reasons for these maneuvers were displayed in a list on the right panel and the currently active maneuver was highlighted. This HMI version also included detected braking of the own and of other vehicles.

Each participant only saw one version of the HMI display (between-subjects factor). The respective HMI, depending on the system's transparency group, was displayed in the center console screen for technical reasons, even though this display was not in the driver's central field of view. In the beginning of the study, participants were made aware of the unusual location of the driving-related information. During the study, people were reminded of the display in the center console, in order to make sure they were aware of it.

Complexity of the situation. The HMI versions were tested in situations that varied in their complexity (within-subjects factor). Situation complexity was understood here as the involvement of other road users that had to be taken into account by the system. Simple, longitudinal traffic situations were mainly free driving or car-following situations, where the HAD vehicle

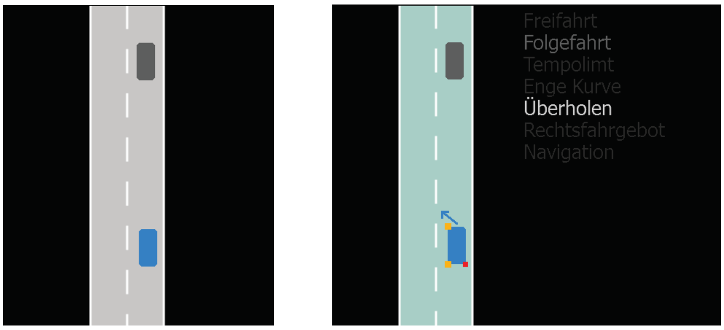


Figure 4.3. HMI versions used in the first user study. HMI version A (left) shows the detected surroundings of the vehicle, represented by the blue rectangle, HMI version B (right) also shows the vehicle's reasoned behavior.

only drove on one lane and did not attempt to change the lane. The more complex situations included also lane change maneuvers and overtaking maneuvers. While in the simple traffic situations the system was mainly required to observe other vehicles in the front, the more complex situations also included adjacent lanes and the back. The specific presence and number of other road users could not be controlled, as the study took place in normal traffic.

Dependent and mediating variables

Subjective and objective data were collected during the first user study, to be able to draw valid and reliable conclusions from the results. A windows tablet with a touch screen was used for the questionnaires to reduce paperwork and facilitate work for both participants and experimenter. The following dependent variables were taken into account.

Factors determining dispositional trust. The factors assumed to influence dispositional trust were assessed by standardized questionnaires: general acceptance of technology (19 items; Karrer et al., 2009), desire to be in control (20-item questionnaire; Burger & Cooper, 1979; Gebhardt & Brosschot, 2002), extraversion (twelve items of the International Personality Item Pool Johnson, 2006), and the tendency to task risks (four items; Desai, 2012). All questionnaires used 5-point rating scales ranging from 1 = 'completely disagree' to 5 = 'completely agree' for answering the items (see Appendix A.1.2, Table A.1) and can be found in Appendix A.1.3. Furthermore, participants were asked to rate the importance of certain trust factors and HMI features.

Situational and learned trust. Trust in the specific system was assessed using the questionnaire of Muir (1989) in the version introduced by Lee and Moray (1992). This questionnaire consisted of four items addressing predictability, dependability, faith, and overall trust in a system, which were to be rated on a 15-point rating scale ranging from 1 = 'very low' to 15 = 'very high' (based on the idea of Heller (1985), see Appendix A.1.2, Table A.2). The items can be found in Appendix A.1.3, Table A.5. Learned trust was assessed by repeated application of the questionnaire after each drive. A similar scale was also used to collect subjective ratings of the perceived performance of the system.

Driving parameters. During the experimental drives, relevant driving data like position, velocity, and lateral as well as longitudinal acceleration of the ego vehicle, parameters of a vehicle in front of the ego vehicle, gas and brake pedal position and the steering wheel angle were stored. Variables collected regarding the automated driving system were the current state of the system, its current maneuvers, and the time until its state changes (time until system is available or time until manual driving is necessary). The data were used to calculate handover and takeover times to an availability indication or a takeover request of the vehicle.

Behavioral data. Video data of the experimental sessions were collected to record any unforeseen or unusual situations as well as the behavior of the participants. The data should enable the identification of reasons for peculiar behavior of participants.

Gaze behavior. Gaze behavior has been used in numerous studies as a measure for attention allocation. To enhance understanding of what information is important for drivers during an automated drive, gaze behavior was recorded to show where drivers put their attention on. The analysis of this behavior was tested as an objective measure of trust. With the Dikablis eye-tracking system (Version 2.5, by company Ergoneers GmbH, 2015), data was collected and processed. The areas of interest included glances to the street, the mirrors, the instrument cluster, and the center displays containing information on the automated driving system (see Figure 4.2).

Distraction. In the last run of the experiment participants were offered to use their smartphone at their own choice (e.g., for surfing, texting). As a measure of distraction, the duration of smartphone use was recorded to assess the driver's voluntary distraction from the driving task. To determine the duration, the periods of time (in s) the driver's gaze was directed to the smartphone were summed up.

Procedure

Starting either in the morning or after midday, the 1.5 hour study began with a brief instruction within the vehicle, explaining the functionality and capabilities of the prototype car and familiarizing the participant with the specific HMI version implemented in the vehicle (see Appendix A.1.1). The research vehicle and its capabilities were described in detail, making sure all participants were aware of the functionality of the vehicle. Participants obtained a brief overview over the different displays and indications and how to interpret them. Participants were then asked to fill out the initial questionnaires on demographics, individual characteristics, and attitudes, as well as paperwork to document that they are responsible for the vehicle and need to obey legal traffic rules. Once all formalities were done, the experimenter explained the further course of the study. When there were no further questions regarding the procedure, the sensors for behavioral data were applied (visible in Figure 4.4).

Figure 4.4. Setup of the first user study inside the automated driving test vehicle.

Participants then started driving manually on urban German roads. The route took subjects on a highway, where the automated driving functionality was available. The route included sections of highways A39 (two-lane) and A2 (three-lane) with normal variation of public traffic. Once the automated driving system was activated with a simultaneous press of two dedicated buttons on the steering wheel, it was able to control speed and distance to other vehicles, and could perform maneuvers if necessary. Participants then were allowed to take their feet away from the pedals and to release the steering wheel. While the automated driving system was active, participants were instructed that the system, when activated, would work highly automated with no action required from the driver. It was explained that nonetheless, they could overrule the system with the brake, the steering wheel or the two dedicated buttons and immediately take back control whenever they felt the need to. Subjects were asked to always engage in automated driving when the function was available, to give them the possibility to experience the functionality as long as possible and be able to give a profound appraisal of the system in the end. At the end of each driving section, drivers got an acoustic and visual indication to resume manual driving. The vehicle informed the driver of an upcoming manual driving section one minute prior to the takeover, and indicated the necessary shift of control again 15 seconds before the end of the automated section.

The drive was divided in four sections that were subdivided by stops at motorway service stations. Each section of the route was 15 to 25 km long, resulting in a driving duration of approximately 10 minutes per run. In the first section of the route, participants had the chance to get acquainted with the HAD system in simple, longitudinal traffic situations. After the first run, the initial ratings for trust and perceived performance were collected. During the following parts of the route, they used the system in either simple or more complex traffic situations in randomized order. Each subsequent stop with a break of 5 minutes was used to rate the previously experienced run. In the last part of the drive, people all engaged the system in complex traffic situations again and were allowed to use their mobile phone as a non-driving-related activity during the highly automated drive. Participants were told to use their phone only when they felt comfortable doing so. While or shortly after using the phone, an unexpected takeover request was started by the experimenter. It was observed how subjects reacted to the unexpected event and takeover times were collected. Every subject experienced each scenario (within-subjects factor), thus each participant had reached the same level of experience after the complete drive. Finally, people were asked to answer a final questionnaire, which included the introduction of the other HMI version in order to compare the two versions directly. The whole procedure can be found in Figure 4.5.

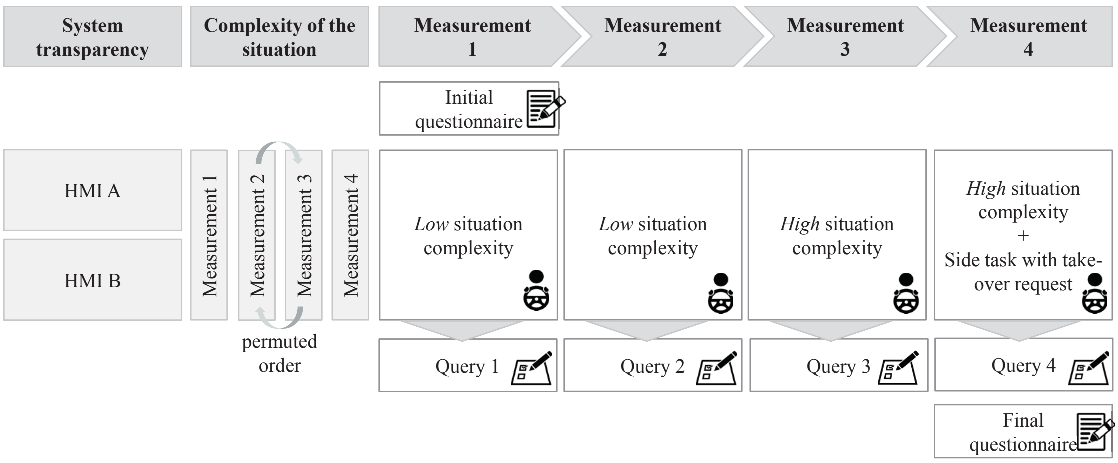


Figure 4.5. Procedure of the first user study.

A trained safety driver with a second gas pedal and brake sat on the passenger seat for safety reasons during the whole drive. The safety driver was familiar with the vehicle and the route of the test drive. The person was introduced to the participants as a technical support, but the intervention possibilities were not mentioned. During the majority of the runs, the automated system worked highly reliable. However, because a real prototype system was used in this study, some participants experienced rare and unanticipated events of low system reliability (e.g., jerks of steering wheel or abrupt braking maneuvers that required a short intervention of the safety driver). The occurrence and number of these events could not be experimentally controlled, but provided an opportunity for a post-hoc assessment of their impact on trust in the system. During the drive, an experimenter sat on the back seat and took notes of unusual situations or behavior of the car, the behavior of the driver, and noted down comments subjects made on the vehicle or the HMI.

4.2.3 Results

To examine the hypotheses postulated beforehand, IBM SPSS Statistics (version 19.0) was used. Figures with statistical data were created with the software R (version 3.4.1). Each of the independent variables was analyzed in detail with inferential statistical methods regarding the assumptions. As the main measure of trust in the system, the mean of the four-item questionnaire of Muir (1989) was used. Due to the four queries of trust gathered with this method, a development of trust could be measured. Overall trust describes the mean of these four queries. Data was analyzed by means of independent samples t-tests or analyses of variances (ANOVA) with repeated measures when appropriate. Before, extreme outliers were removed from the analyses. The Shapiro-Wilk test was used to check the distribution of the dependent variables for statistical normality. In case the requirements for a t-test or an analysis of variance were not met, non-parametric tests (Mann-Whitney-U or *Welch's F*-test) are reported. The two main factors as well as factors potentially influencing dispositional trust in the system, like personality characteristics and attitudes, and factors affecting learned trust were taken into account. Significant results of the collected data are presented in detail. Significance level was defined as $\alpha = .05$. An overview over the main variables and their correlations is given in Appendix A.2.1, Table A.14.

Hypothesis 1: Dispositional trust

In the first hypothesis it was assumed that initial trust in an automated vehicle would be affected by the user's personality and attitudes toward technology. This hypothesis was tested by contrasting initial trust ratings of participants with comparatively high vs. low scores on the different personality and attitude scales (median split) by means of independent samples t-tests.

Acceptance of technology. An important covariate assumed to have an impact on initial trust in an automated system was acceptance of technology. A significant difference in initial trust ratings was found for acceptance of technology, with the group with low acceptance showing less trust ($M = 9.96$, $SD = 2.69$) than the group with high acceptance ($M = 12.29$, $SD = 2.25$), $t(24) = 2.06$, $p = .050$, $d = 0.81$ (one case was identified as an extreme outlier and was thus eliminated from the analysis). The difference between the groups can be seen in Figure 4.6 (left). A positive correlation between initial trust ratings and technical affinity verified this result (pearson correlation coefficient $r(25) = .44$, $p = .023$), explaining approximately 18% of the total variance of initial trust ratings.

Desire for control. Desire to control was found to be another important variable affecting initial trust levels in automated driving systems. Distinguishing between a group of participants having a low desire for control and participants with a high desire for control, a significant difference in initial trust ratings could be found for desire for control with participants with a high desire for control exhibiting less trust in the system ($M = 9.55$, $SD = 3.29$) than participants with a low desire for control ($M = 11.93$, $SD = 2.11$), $t(26) = 2.27$, $p = .032$, $d = 0.86$. A visualization of this effect can be found in Figure 4.6 (right). No correlation between desire for control and initial trust ratings were found.

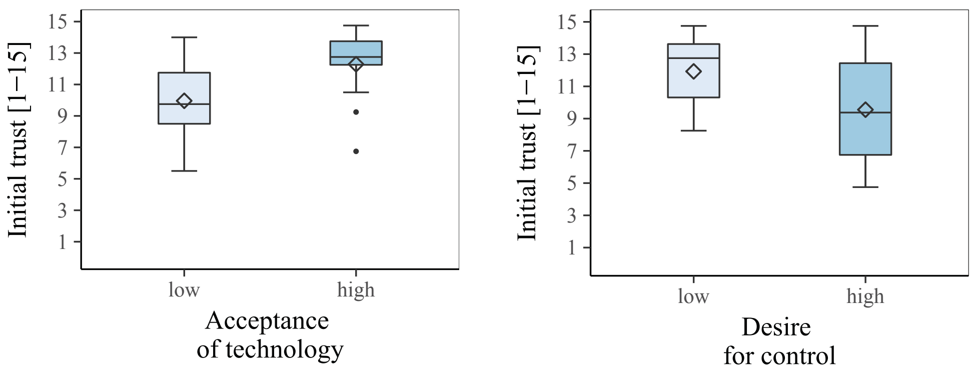


Figure 4.6. Initial trust ratings for the covariates acceptance of technology (left) and desire for control (right). The boxplots show the results of the subjective initial trust ratings of participants with comparatively high vs. low scores. Error bars represent the standard deviation.

Statistical analyses were also conducted for the factors assessment as a driver, current stress level, extraversion, and the tendency to take risks. No significant differences emerged for either of these factors between pooled low- and high-level groups for the initial trust ratings. A small difference could be found for the investigated factor self-efficacy, with the low self-efficacy group stating to initially have lower trust in the system ($M = 9.13$, $SD = 2.86$) than the high self-efficacy group ($M = 11.67$, $SD = 3.15$). The Mann-Whitney-U test was used to analyze the difference between the groups, as initial trust ratings were not normally distributed in one group (Shapiro-Wilk test, $p = .035$). A significant difference was found with this more robust test, $U = 29.50$, $p = .043$, $d = 0.84$.

Regression of trust. Three of the tested covariates were found to be related to initial trust in the system and could potentially be used as predictors for the dependent variable initial trust ratings. Complementary to the foregoing analyses, a stepwise multiple regression analysis with the variables acceptance of technology, desire to be in control, and self-efficacy was used to reveal how much of the criterion's variance can be explained by these covariates. The regression coefficient acceptance of technology was able to account for approximately 16% of the variance of initial trust ratings ($R^2 = .16$). This explained percentage of variance is significant compared to the total existing variance, $F(1,25) = 5.87$, $p = .023$. As the p-value is smaller than .05, the chosen factor predicted the dependent variable significantly better than would be expected by chance. Acceptance of technology was the strongest predictor for initial trust in the automated driving system ($\beta = .44$, $t(26) = 2.42$, $p = .023$), while desire to be in control only explained a smaller percentage of variance (explanation of additional 12% of variance if added to the model, with $\beta = -.41, t(26) = -2.05$, $p = .051$). Self-efficacy could not improve the prediction quality significantly when added to the model. The relationship between the two main covariates and initial trust in an automated vehicle are shown in Figure 4.7.

Besides the personality characteristics, a closer look was also taken at the influence of demographic factors that could potentially influence trust, such as gender and age. While no significant differences were found in subjective initial trust ratings for men and women (using an independent *t*-test), a significant effect was found for different age groups.

Trust and Age. Two age groups were classified according to the median of 33.50 years to distinguish between younger and older people of the sample (median split). The independent *t*-test revealed an effect of the factor age group on initial trust in the automated driving system, ($t(26) = 3.03$, $p = .005$, $d = 1.15$). The analysis thus indicated that younger people have a significantly higher initial level of trust compared to the older age group. A negative correlation of the factor age with initial trust ratings was found (pearson correlation coefficient $r(26) = -.51$,

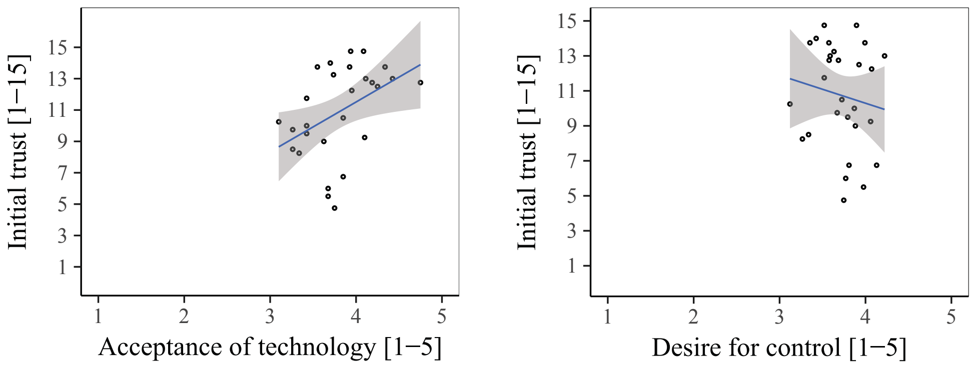


Figure 4.7. Relationship between initial trust ratings and the covariates acceptance of technology (left) and desire for control (right).

$p = .005$). Thus, it can be concluded that approximately 26% of the total variance of initial trust ratings are accounted for by the factor age. No correlations were found between the covariates age and acceptance of technology, desire for control, or self-efficacy. When the covariate age was added to the multiple regression analysis presented above, the regression coefficients acceptance of technology, desire for control, and age were able to account for approximately 38% of the variance ($R^2 = .38$), $F(3,23) = 6.29$, $p = .003$. The factor age could explain additional 14% of variance when added to the model, with $\beta = -.39$, $t(26) = -2.42$, $p = .024$. The left side of Figure 4.8 shows a boxplot with the results regarding subjective trust ratings of the two age groups. On the right, the regression for initial trust ratings is depicted.

Trust and other factors. In addition to the analysis of their subjective ratings for trust in the automated driving system, participants were also asked to assess the influence of certain factors regarding trust in automation. The relevance for trust was rated on the 15-point scale by Heller (1985) (see Appendix A.1.2, Table A.2). The result of this question can be found in Figure 4.9, showing a great importance of reliability and technical capabilities of a system for the development of trust in it.

Hypothesis 2: Situational trust

Hypothesis 2 postulated that situational trust will be lower in complex driving situations. The HAD system was able to handle situations of different complexity, e.g., car following situations as well as takeover maneuvers. However, drivers did not know the capabilities of the prototype vehicle, as they had no previous experience with the HAD system. In the first run, drivers got

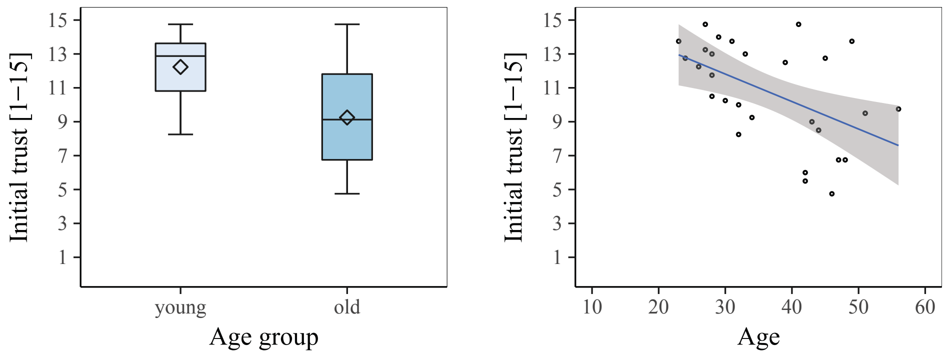


Figure 4.8. Initial trust ratings for the covariate age (left) and the relationship between initial trust ratings and the covariate age (right). The boxplot shows the initial trust ratings of the two age groups. Error bars represent the standard deviation.

acquainted with the system. The second and third run included either easy or more complex driving situations (in randomized order). When comparing nervousness ratings for the two situation complexity levels, a tendential effect could be found with a paired samples t-test, $t(27) = 2.04$, $p = .052$, $d = 0.39$. Drivers were slightly more nervous during more complex driving tasks ($M = 6.21$, $SD = 3.55$) compared to easy driving tasks ($M = 5.00$, $SD = 3.15$). When comparing trust ratings, another tendency was found, $t(27) = 2.00$, $p = .056$, $d = 0.38$. This effect can be seen in Figure 4.10. As postulated, participants trusted the HAD system slightly more during easy driving situations ($M = 11.29$, $SD = 2.74$) than during more complex driving situations ($M = 10.72$, $SD = 2.75$).

The complete statistical model with the factors situation complexity and system transparency showed a similar result. The repeated measures analysis showed a tendential effect of the factor situation complexity, $F(1,26) = 3.86$, $p = .060$, $\eta_p^2 = .13$, while no effect of system transparency was found, $F(1,26) = 1.46$, $p = .238$.

Hypothesis 3: Learned trust

In hypothesis 3, trust was expected to develop depending on the experiences made while interacting with the HAD system. Specifically, it was expected that trust ratings would increase over time and with accumulating experience. In addition, also the transparency of the automation, operationally defined by the different HMI versions, was assumed to influence trust.

Experience and system transparency. As a first step of the overall analysis, data was analyzed by a 2 (system transparency) x 4 (runs) repeated measures analysis of variance. Neither

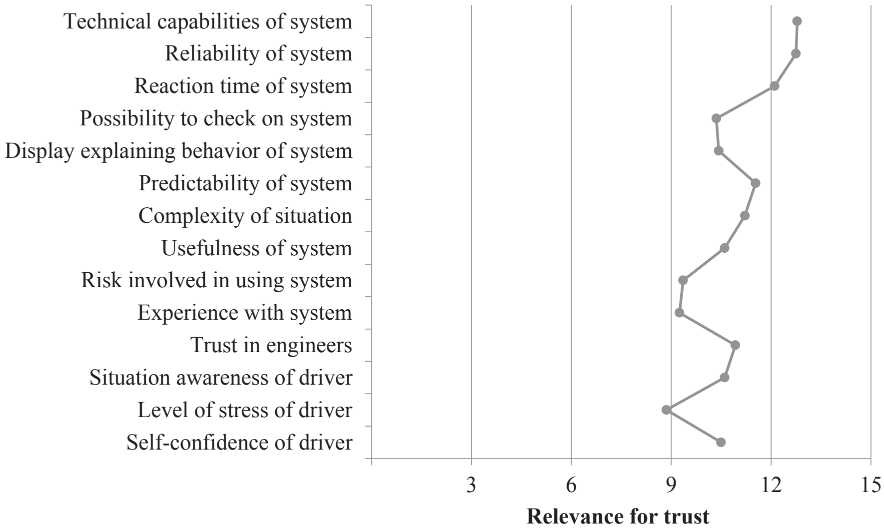


Figure 4.9. Subjective relevance ratings for trust in an automated system.

the main effects of system transparency, $F(1,26) = 0.50$, $p = .484$, and of runs, $F(3,78) = 1.17$, $p = .326$, nor the transparency x runs interaction effect, $F(3,78) = 2.05$, $p = .114$, became significant. No effect of rising experience with the system over the course of time could be found, even though ratings for trust were lower in the beginning of the study. Interacting with an automated driving system with a certain level of transparency did not influence the trust assessment in this study. Even though system transparency did not influence trust ratings, when analyzing which HMI version people preferred in direct comparison, a clear majority of participants (25 of 28) was in favor of the more comprehensive HMI B, $\chi^2(1, N = 28) = 17.29$, $p < .001$ (see Figure 4.11).

When taking a look at the importance of HMI features rated in the final questionnaire, one can see that especially the instructions for using the system were of high importance, as well as the indication of speed limits and information about upcoming construction zones or traffic alerts. Subjects were asked to provide an assessment of importance for short system use (first-time use) and longer system use (approximately half a year). The deviation between the two profiles is small, indicating that the importance of the HMI features does not change over the course of interaction with the system (see Figure 4.12).

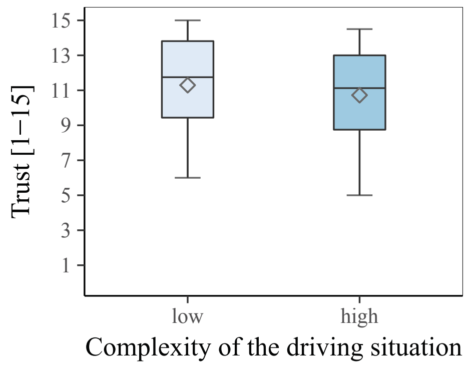


Figure 4.10. Subjective trust ratings for the factor situation complexity. The boxplot shows the results of the two levels of low (light blue) and high (dark blue) situation complexity. Error bars represent the standard deviation.

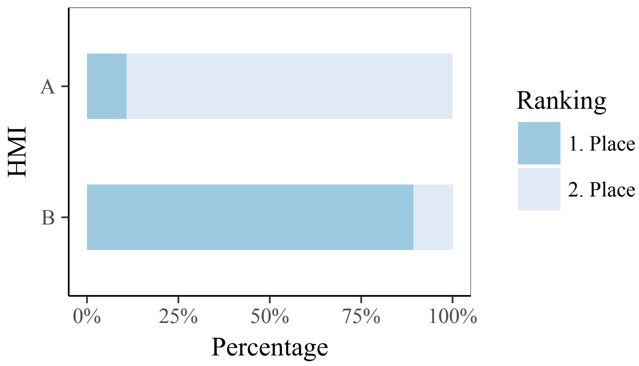


Figure 4.11. Results for subjective HMI preference in the first user study. The bar plot shows the percentage distribution of ratings for the two HMI concepts.

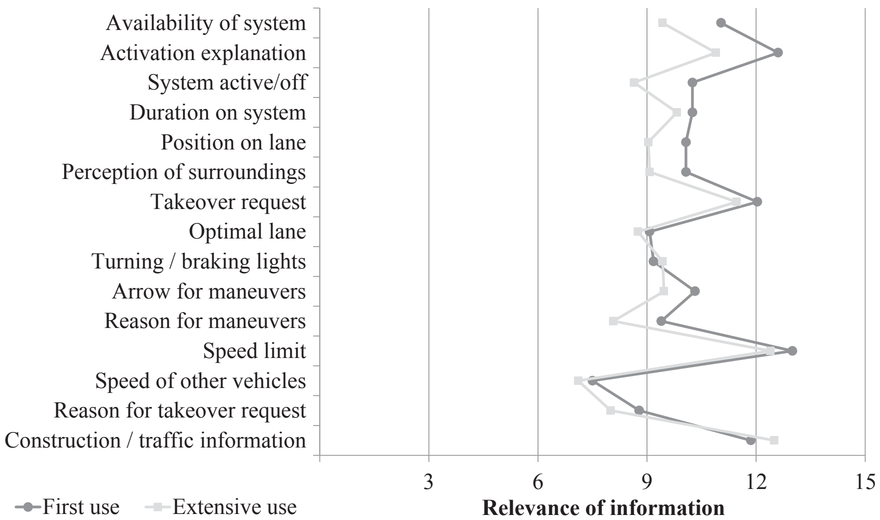


Figure 4.12. Results for subjective importance of information provided by the automated driving system.

Perceived performance. As a subsequent analysis regarding system characteristics, data of the HMI groups was aggregated to analyze the impact of the covariate perceived performance of the system on the development of trust. The perceived performance of the automated driving system was collected for each section of the drive in order to assess the impact on trust development in the highly automated vehicle. Two groups differing in the mean level of subjectively perceived performance of the automated driving system were defined by a median split. The repeated measures analysis of variance with the additional factor perceived performance showed that trust ratings differed significantly between those two groups, $F(1,26) = 36.23$, $p < .001$, $\eta_p^2 = .58$, confirming the importance of perceived system performance. Congruently, a positive correlation was found for all different sections of the drive. Pearson correlations between trust and performance ratings indicated that trust ratings got determined by perceived system performance in each measurement (pearson correlation coefficient run 1: $r(26) = .64$, $p < .001$; run 2: $r(26) = .77$, $p < .001$; run 3: $r(26) = .58$, $p = .001$; run 4: $r(26) = .80$, $p < .001$). The chronological sequence of trust ratings depending on perceived performance groups is presented in Figure 4.13. In each run, lower trust was reported by participants that perceived the system's performance as low.

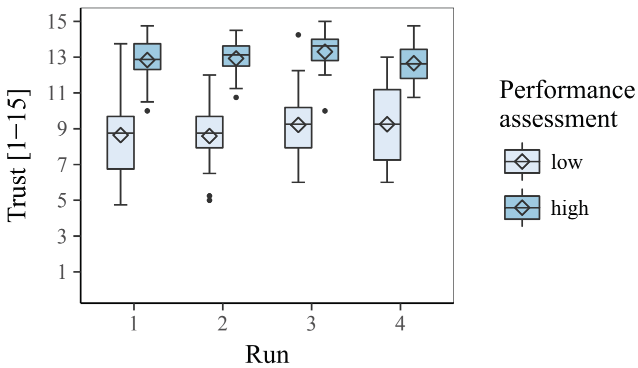


Figure 4.13. Subjective trust ratings for the factor performance. The boxplot shows the resulting trust ratings of the group with low performance ratings (light blue) and the group with high performance ratings (dark blue) over the course of the four queries. Error bars represent the standard deviation.

Trust and system boundaries. In a third analysis, system boundaries were scrutinized in detail. As has been described above, participants experienced different numbers of unanticipated system reactions while driving, depending on environmental influences which were not controllable in advance. As the study was a real driving study and did not take place in a simulated environment, environmental conditions could not be controlled entirely. One major aspect that needs to be taken into account was the reliability of the system when system boundaries were met. As the concept car is a research platform, occasions of low system reliability were still possible (justifying the safety driver as a last redundancy). However, even though the driver (or the safety driver) reacted promptly to a sudden jerk of the steering wheel or abrupt braking maneuvers of the vehicle, these low reliability events could not be hidden from the participant. Thus, the undesired system reactions needed to be considered in the subsequent analysis in case they affected the data.

To investigate how the experience of such events would impact trust in the system, an exploratory analysis was performed by contrasting trust ratings of participants who had experienced no (n = 13), one (n = 11), two (n = 3), or three events (n = 1) of low system reliability during the four runs. The mean trust ratings of these subgroups are shown in Figure 4.14. As becomes evident, the experience of two such events diminished trust considerably, while one such event seemed to remain without any consequence on trust. In a post-hoc comparison between a group that experienced not more than one event and a group that experienced two or more events, this observation was confirmed, $U = 17.00$, $p = .042$, $d = 1.26$ (the Mann-Whitney-U test was used due to the different sizes of the groups).

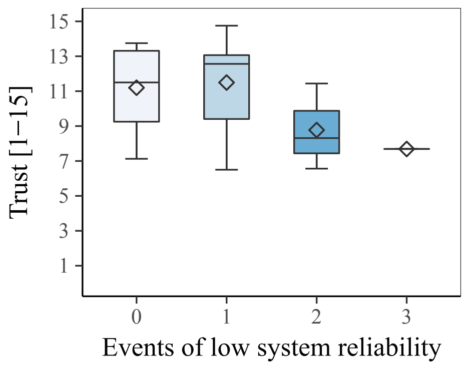


Figure 4.14. Subjective trust ratings depending on the number of low reliability events. The boxplot shows the trust ratings in relation to the number of low system reliability events participants experienced. Error bars represent the standard deviation.

Whenever participants noticed that the safety driver was actually able to control the car in the event of low system reliability, a note was taken in order to analyze these cases separately. Of all participants, six drivers realized the intervention possibilities their passenger had. Due to the unequal group sizes, trust ratings were compared with *Welch's t*-test (Levene-Test $p = .002$). Results show that the subjective ratings for trust in the system are indeed confounded with people's trust in the human safety driver. People who realized that a safety driver was on board rated their system trust higher than participants who did not realize the other human as a fallback solution, *Welch's* $t(26) = 4.38$, $p < .001$, $d = 1.42$.

Hypothesis 4: Trust and driver behavior

In the fourth hypothesis it was proposed that drivers' trust in the system would determine attention allocation. Drivers with a comparatively low trust were expected to allocate more attention to the driving situation than drivers with a comparatively high trust in the system. To investigate this hypothesis, gaze behavior, engagement in smartphone use while driving, and takeover times were analyzed depending on the trust level (median split).

Attention allocation. A high level of trust in a system can lead to complacent behavior and reliance on the automation used for the task at hand. When it comes to driving highly automated on the highway, the first thing to do when not being forced to pay attention to traffic any longer is to look around freely and let the eyes wander wherever they want. Whether the driver dares to do that or is still observing traffic and the vehicle is possibly a question of his level of trust in the system. It is assumed that gaze data can be an objective measure to determine the level of trust the driver has. In this work, percentage gaze distribution (attention ratio) on the AOIs (street, instrument cluster, center console displays, and mirrors) was gathered over all runs to identify differences in gaze behavior. Gazes at the mirrors were interpreted as checking on the system. Also, monitoring other relevant aspects of the driving task like instruments and displays was construed as supervising behavior. When examining gaze behavior of the group with a higher trust level compared to the group with a lower trust level (median split), a significant difference regarding the attention on the instrument cluster was observed, *Welch's* $F(1,15) = 5.62$, $p = .032$, $\eta_p^2 = 0.22$. Participants that were unsure about the system's capabilities or behavior searched for information in the instrument cluster, where other driving related information was displayed. They had lower trust in the system and checked the instrument cluster more ($M = 14.04\%$, $SD = 8.92\%$) than participants with high trust ($M = 6.90\%$, $SD = 4.51\%$). No significant differences were found for the AOIs street, center console, and mirrors. A negative correlation between overall trust ratings and percentage of gazes to the instrument cluster ($r(20) = -.55$, $p = .009$) indicated the same result. The results can be seen in Figure 4.15 that visualizes the differences in percentage distribution of gazes between the two groups regarding specific AOIs.

As an addition to the analysis of the trust groups, the two different HMI versions were compared. An analysis of variance was conducted, showing a significant difference between the groups A and B regarding the percentage of gazes on the street, $F(1,20) = 11.53$, $p = .003$, $\eta_p^2 = 0.34$, and on the HMI screen in the center console, $F(1,20) = 4.40$, $p = .049$, $\eta_p^2 = 0.18$.

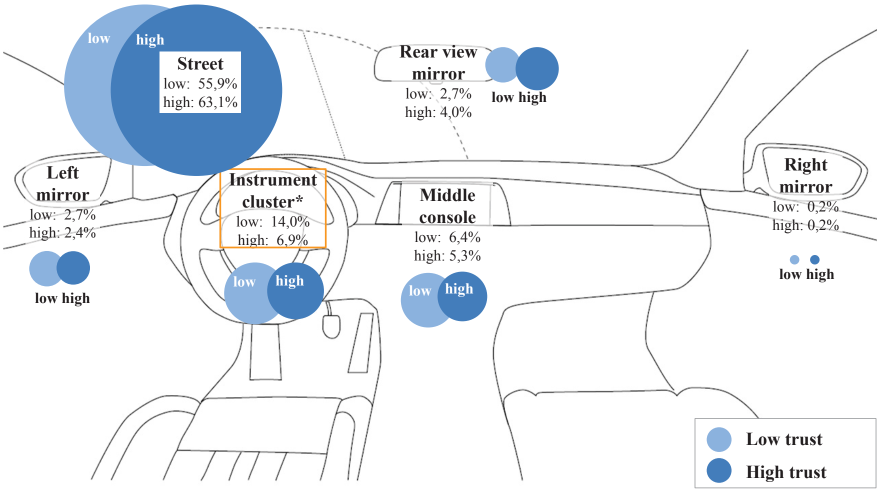


Figure 4.15. Percentage of gazes (attention ratio) depending on low trust (light blue) or high trust (dark blue) in the automated driving system. The size of the circles is relative to the percentage of gazes in this area.

The results are visualized in Figure 4.16, with bigger circles representing a higher percentage of gazes and smaller circles indicating less attention in this area. Significant differences are marked with an asterisk and an orange box. Unsurprisingly, the more comprehensive HMI B got more attention, as it contained more information that needed to be processed.

Use of smartphone. For the final section of the study, participants were allowed to use their smartphones while automated driving was active. Over 90% of all participants tried using their smartphone while driving. Two drivers did not dare to use their smartphone, the rationale being legal aspects (not being allowed to use a smartphone while driving during normal circumstances). Instead, they turned around to talk to the experimenter on the back seat during the drive. Participants who trusted the automated driving system took advantage of this opportunity and completely engaged in smartphone use. However, participants not feeling that sure about the automated system rather checked the surroundings from time to time, especially during maneuvers like lane changes, where lateral acceleration was higher than during the normal straight drive. They only used their smartphone for brief periods. Data was pooled to compare a group of participants with low trust in the system with a group of participants with high trust (median split). Percentages of gazes directed to the smartphone were gathered to contrast the mean duration of smartphone use of the two trust groups. An independent samples t-test of this data revealed a significant difference, $t(23) = 2.11$, $p = .046$, $d = 0.83$, indicating that participants with a high level of trust used the smartphone significantly longer ($M = 169.42$ s, $SD = 75.73$ s)

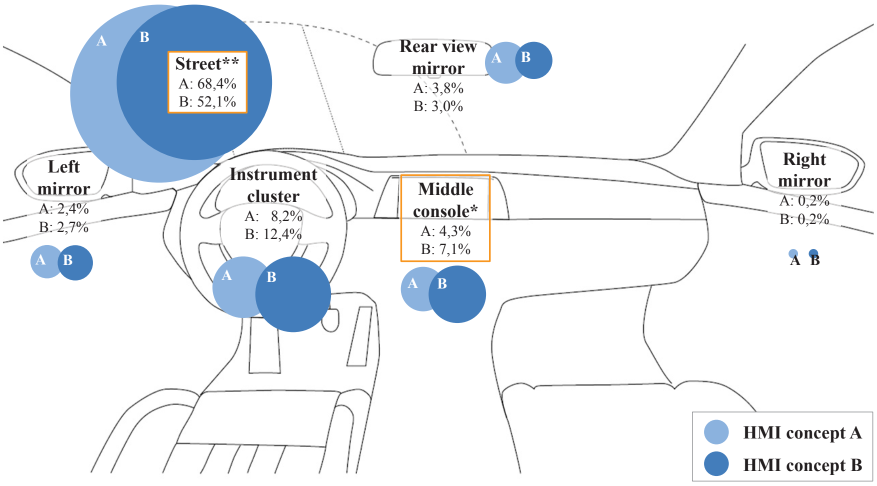


Figure 4.16. Percentage of gazes (attention ratio) depending on HMI A (light blue) or HMI B (dark blue) during the highly automated drive. The size of the circles is relative to the percentage of gazes in this area.

than participants not trusting the system ($M = 119.92$ s, $SD = 36.46$ s, shown in Figure 4.17). A positive correlation between overall trust ratings and duration of smartphone use was also found ($r(23) = .49$, $p = .014$).

Takeover time. Another objective indication for a high trust and reliance on the system was assumed to be the reaction time to a takeover request of the vehicle. Whether participants take longer because they are busy with something else or whether they just feel they can take their time until they are ready to take back control—either way it suggests that they feel safe while in automated driving mode and trust the system to work. The assumption is that the faster their takeover time is, the stronger is their urge to take back control. In addition, also the time until activation from the moment the system is offered was analyzed. Looking at the development of takeover times in Figure 4.18, it could be observed on a descriptive level that participants activated the system faster and took more time until taking back control over the course of time. Only in the last run, when participants had experienced a sudden takeover request before, the takeover time was quicker again.

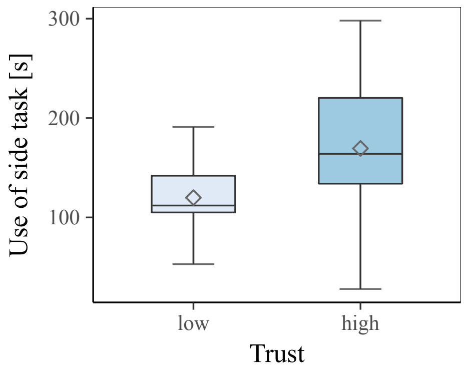


Figure 4.17. Duration of use of the non-driving-related activity. The boxplot shows the duration of side task use of participants with comparatively high vs. low trust. Error bars represent the standard deviation.

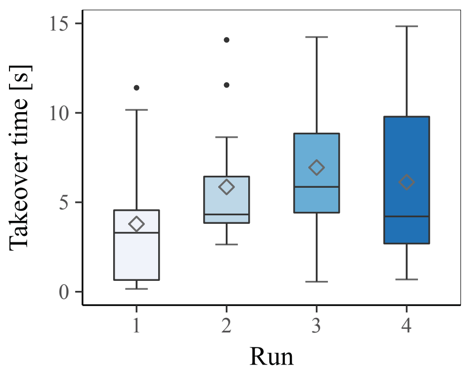


Figure 4.18. Handover time of the drivers over the course of the four runs. The boxplot shows the takeover time over the course of the four runs. Error bars represent the standard deviation.

An analysis of mean handover and takeover times depending on the level of trust in the system (median split) was conducted to find out whether participants' interaction with the system changes with rising trust. Even though the mean reaction times of the two trust groups to a takeover request differed in the assumed direction, with a mean reaction time of 3.88 s ($SD = 3.09$ s) for the low trust group and 5.77 s ($SD = 3.09$ s) for the high trust group (see Figure 4.19, right), this effect did not constitute a significant difference. The time until activating the system after the indication was longer for participants with higher trust in the system ($M = 6.21$ s, $SD = 2.05$ s) than for participants with lower trust ($M = 4.10$ s, $SD = 1.90$ s). An independent t-test with the level of trust as the differentiating factor showed a significant difference in the time until activation, $t(26) = 2.83$, $p = .009$, $d = 1.07$, as Figure 4.19 (left) shows.

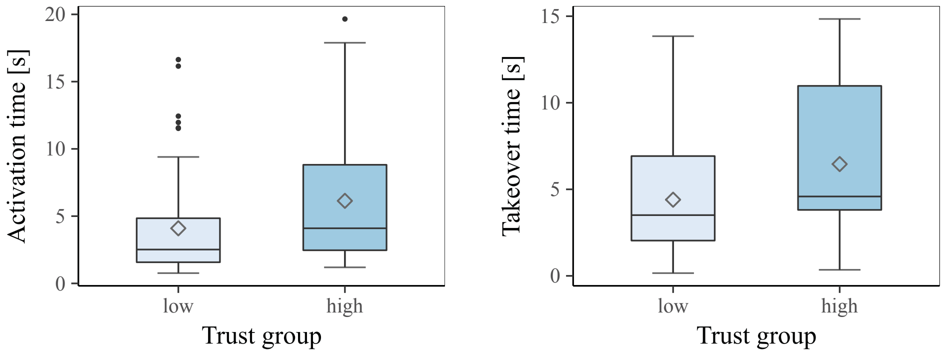


Figure 4.19. Handover time until activating the automated driving system after the offer (left) and takeover time until deactivating the system after the takeover request (right) depending on a low level (light blue) or a high level (dark blue) of trust in the automated driving system. Error bars represent the standard deviation.

4.2.4 Implications

The objective of the first user study was to evaluate the influence of individual and system characteristics on trust in automated driving and to enhance knowledge about the development and consequences of trust in HAD. In four runs of driving automated on the highway with a duration of 10 minutes each, 28 participants experienced dedicated HMI concepts (*system transparency*) under simple and more complex driving conditions (*situation complexity*). Personality characteristics were also taken into consideration in the analysis. The study was able to confirm former research results in a real driving environment, and expand existing knowledge about factors influencing trust in automated driving by adding new results based on a driving study in a real environment.

Findings. The first hypotheses stated that initial trust in an automated driving system would be determined by global personality characteristics like extraversion as well as specific attitudes. Several personality characteristics were assumed to have an influence on dispositional trust in an automated driving system. The results suggest that particularly individual differences with respect to technical acceptance and desire for control influence initial trust in this novel technology. The hypotheses 1a (acceptance of technology) and 1d (desire for control) can thus be accepted. People having a low technical affinity or a high desire for control will trust an automated driving system less in the beginning. In addition, the demographic factor age was found to predict dispositional trust in an automated driving system. Younger participants tended to have more trust in the system, independent of other personality factors. However, other factors that have been observed in laboratory studies to impact trust in automation were not found to have an influence in this real driving study (e.g., extraversion, Merritt and Ilgen (2008)). It appears that at least in this context, specific attitudes and characteristics of personality were more important than global personality factors for determining trust in the automated driving system. All in all, hypothesis 1 can be partly confirmed. Particularly, age and acceptance of technology were found to be good predictors for initial trust in an automated driving system. When asked about other factors apart from personality, people furthermore state that reliability and technical competence are seen as the most important system characteristics for developing trust in the system. The complexity of the situation was found to have a tendential effect on trust as well, as was postulated in the second hypothesis. Participants had slightly more trust in the HAD system when the driving situations were easy. Trust development over the course of actively experiencing the system was mostly influenced by the perceived performance of the automated driving system. This finding confirms many other studies (De Vries, 2004; Kazi et al., 2007; Lee & Moray, 1992, 1994; Merritt et al., 2013; Muir, 1994; Muir & Moray, 1996) that point to the importance of system reliability and perceived technical competence. This study's analysis of trust ratings in the context of automated driving confirms that low perceived performance of the vehicle leads to less trust of the driver. The correlation between perceived performance and trust was high for each run. A particularly interesting finding of this study is the effect of events of low system reliability on trust. Meeting boundaries of the system was followed by a decrease in trust. Surprisingly, only the experience of repeated events led to decreases in trust. One single event did not diminish trust significantly, contrasting results from the laboratory which often reported significant reductions of trust already induced by single automation failures with only a slow recovery from it (Lee & Moray, 1992; Ma, 2005; Muir & Moray, 1996). Overall, hypothesis 3a cannot be confirmed, as trust did not rise with growing experience, but was diminished when experiencing more than one unintended system reaction. The observation of unintended automation behavior had an influence on trust ratings. As the study setting was unable to control for these events, participants experienced a different number of unpredictable system reactions, depending on weather conditions, traffic, and other environmental influences the automated driving system needed to cope with. The observations show that experiencing one event of low system reliability does not have an impact on trust as great as experiencing more than one such event. Comments of the participants indicate that they are looking for an explanation for low reliability events. One event may be explained by certain environmental circumstances, and may thus be attributed externally. Several events can show inconsistent patterns that can no longer be attributed to one external reason. They might rather be attributed to the system, resulting in less predictability and in uncertainty for the driver. Reasons for low reliability given by the vehicle could help to preserve trust further. Contrary to the expectation of hypothesis 3b, the different HMI concepts varying the transparency of the system did not influence the development of trust. This result is similar to Kleen et al. (2014), who used an 'Active Frame' concept for system feedback in an automated driving study, but did not find an effect on system trust. One possible explanation may be that the factor's variation in the present study was not strong enough to make a difference. However, although there was no significant effect of the main factor system transparency on trust, nearly 90% of the participants still preferred the more comprehensive HMI concept in direct comparison of the two alternatives.

Finally, the fourth hypothesis addressing behavioral indicators of trust mostly got support by the data. Drivers trusting the system more were also more willing to use their smartphone, distracting themselves from the driving scene. They used their smartphone more extensively by making a call or texting, and were thus distracted from the driving situation for a longer time period instead of observing the vehicle's drive. Also, the activation of the system took people with more trust longer, indicating that people took their time when interacting with the system instead of hectically reacting to a system request. Takeover times were not affected, against the initial assumption based on findings of Pradhan et al. (2015). Furthermore, participants not trusting the system looked at the instrument cluster more than trusting participants. Although the instrument cluster did not show system-related information, these drivers seemed to have expected more information about the drive there. The results need to be interpreted with caution, however, as some of the variables were not normally distributed. The insights are important because links between trust and behavior have rarely been found in research on trust in automation (one exception being Hergeth et al., 2016) and are therefore controversially discussed. The results at hand suggest that trust is an important variable determining driver's behavior, relevant for attention allocation as well as the use of a non-driving-related activity. Knowing this, the observation of these variables in turn can help to objectively assess trust in an automated driving system.

The comprehensive HMI got more attention than the simpler HMI version. The position of the detailed system information seemed to be not ideal. The inconvenient position of the information might be a reason for the unaffected trust ratings. A more central location than the center console (ideally in the visual axis in front of the driver) should be preferred, because people not trusting the system are looking for further information in the instrument cluster, where driving related information is normally displayed. Thus, to enhance trust in the system, information should be displayed where people are looking for information. Furthermore, an effect of system's transparency might have been masked by other, more pronounced effects, like the effect of system performance on trust.

Limitations. Limitations of the first study that need to be considered include its limited internal validity due to its setting in real traffic. As the study did not take place in a fully controlled environment, some factors cannot be ruled out to have affected the results. For example, it could not be controlled for the presence and volume of traffic, roadway and weather conditions, and thus events of low system reliability, making a structured analysis of this factor's impact difficult. In addition, the novelty of the situation and the large amount of information may have led to an intensified concentration on performance, masking (small) effects of HMI and changes over time. In addition, the setting in a prototype vehicle and in real traffic required the presence of a safety driver. Although questions in the questionnaire always inquired specifically regarding trust in the automated driving system and not in any person related to the study, an effect of the safety driver was found. Thus, results were confounded with non-intended effects of trust in a human passenger. However, for safety reasons the human safety driver was crucial, and it was not possible to prevent this effect entirely.

Regarding the HMI concept, the location in the center console was found to be inconvenient. Participants were looking for information in front of them, thus driving-related information should be displayed there. The HMI concept was thus not displayed as prominent as was intended. Furthermore, the experimental driving duration of not more than one hour might have been too short to establish learned trust in the system. However, these compromises had to be taken to extend human factors research related to automated driving into real field settings.

Another study limitation was the relatively small sample size of 28 participants. The limited availability of the prototype vehicle did not allow for a larger study setup or a larger sample size. Also, participants were all employees of the Volkswagen Aktiengesellschaft. The sample therefore may be not completely representative, and results need to be interpreted cautiously.

Conclusion. To conclude, the presented study provides research results on trust development in automated driving with new insights due to its setting in a real driving environment. The results of the user study were able to confirm that certain personality characteristics, like desire for control, and certain personal attitudes, like technical acceptance, have an influence on trust in automated vehicles. The most important factor forming learned trust during the interaction with an automated vehicle turned out to be the perceived performance of the system. More precisely, results suggest that single events of low automation reliability might be tolerated, but experiencing more than one situation of that kind in close succession can diminish trust significantly. The finding that the level of trust is directly reflected in drivers' attention allocation and smartphone use while driving automated proves the practical relevance of trust research in the context of automated driving. Moreover, it suggests that trust in automated driving can be indirectly assessed by observing specific characteristics of driver behavior (e.g., gaze behavior and use of a secondary task) as objective measures. Building on the results presented here, subsequent user studies should elaborate the interaction of different factors determining trust in automated driving systems further and identify the information drivers need to understand the capabilities and boundaries of an automated driving system and develop appropriate trust in it.

New research questions emerge from these insights. It is still unclear whether a detailed HMI concept can help drivers to form greater trust in a system, for example in case of low system reliability. The more detailed HMI used in the study was preferred, but it did not enhance trust in the system compared to lower system transparency. To support the development of trust, it may be necessary to place the information in a more central position, and moreover to inform the drivers in a more detailed and predictive way about how the system behaves. This way, system's transparency might have the potential of helping trust to develop, as research suggests (Beggiato & Krems, 2013; Cai & Lin, 2010; Helldin et al., 2013; Lee & See, 2004; McGuirl & Sarter, 2003; Verberne et al., 2012; Ye & Johnson, 1995, see also Section 3.1). Also, events of low system reliability were not examined in a controlled environment, thus hindering a structured analysis of these events. A study with varying system reliability under controlled conditions seems necessary to confirm the results of this field study. The restrictions of the first user study are addressed in the second user study that is described in the next section.

4.3 Study 2: The importance of system reliability

The second user study was designed to answer the research questions raised by the first study. It was conducted to find out how unanticipated system reactions can influence trust development in the context of HAD, as in the first user study perceived system reliability and the number of events of low system reliability had been shown to influence trust immensely. Furthermore, it was investigated in the study if a dedicated HMI concept can support driver's trust in the system and can guide it during such events. In addition, the influence of experience with the automated driving system was also examined.

4.3.1 Hypotheses

Based on the theoretical background reported so far and the results of the first user study presented above, the following research findings were expected.

To enhance the driver's knowledge about system behavior, capabilities, and boundaries and through this to promote trust, different approaches can be pursued. Information can be given beforehand (e.g., in the form of a manual or training) or can be developed through the interaction with the system (e.g., by using an HMI to give feedback out of the system). Both ways were investigated in the second user study. It was suggested before that familiarization and experience with the system may influence trust in it (Koustanai, Cavallo, Delhomme, & Mas, 2012). The first possible influence tested in the user study aims at the development of a comprehensive mental model before the actual use of the system. Preceding knowledge about the automated driving system (in this case due to prior explanations about the system) resulting in an adequate mental model defines whether a person is an inexperienced or an experienced user. The effect of the level of experience with the system shall be clarified by modifying the preliminary information about the automated driving system.

It was expected that the *level of experience* with the system influences the development of trust in an automated driving system.

> *Hypothesis 1: Trust in an automated driving system is influenced by prior knowledge about the system (experience).*

The importance of a fully developed mental model for the reduction of uncertainty of a situation has already been highlighted in Section 2.3.1. Experience was thus suggested to support a higher level of trust in a system compared to inexperienced system use.

> *1a) Trust in an automated driving system will be positively influenced by prior experience with the system.*

One objective of HAD is to add to driving comfort. However, the first pre-study with the prototype automated vehicle had indicated a higher level of stress during automated driving compared to manual control (see Section 3.3.1). The low level of experience with the system might have been one reason for the high level of stress during these drives. What remains unclear is whether the stress level will eventually drop when the driver gets to know the system. For the second user study, drivers of the experimental group with experience were assumed to experience less stress during the drives than the control group. It was furthermore expected that participants with experience will need less information during the automated drive than the control group without experience with the system.

> *1b) The level of stress will be lower when having prior experience with the system.*
>
> *1c) Experienced users of the automated driving system will need less information about the system's behavior and will, over time, rate detailed information as less useful and relevant than inexperienced users.*

To enhance knowledge about the system during the interaction with the automated driving system itself, detailed information integrated into the system with the help of a dedicated HMI concept was expected to give users the chance to develop the mental model of the system even further. In the first user study, this factor did not have an effect on trust in the system, however it remained unclear if in certain situations (e.g., when unanticipated system reactions occur) information of an interface can help to interpret system behavior better and to maintain trust in the system.

> *Hypothesis 2: Trust is affected by providing detailed information about the system's state and behavior to the driver compared to less detailed status information (system transparency).*

The factor *system transparency* was assumed to have an influence on initial trust ratings as well as on trust ratings after an event of low system reliability. No such effect had been found in the first user study. In this second study, the HMI was improved graphically and presented right in front of the driver to make it more noticeable. Furthermore, the influence during situations of low system reliability was of interest in this study (see hypothesis 3c).

- 2a) *Detailed information about the system's behavior will help to create trust in the automated driving system.*
- 2b) *The level of stress will be lower when providing more information about the system during the drive.*
- 2c) *Detailed information about the system's behavior will be rated more useful than less detailed status information (even though the driver is not in charge anymore).*

The first user study indicated that perceived system performance is especially relevant for the development of trust in an automated driving system. Trust has thus been proven to depend on the experience of high or low system reliability. However, results indicated that more than one event of low system reliability is needed to diminish trust substantially. The second user study varied the reliability of the system to research this effect in more detail (e.g., through unexpected maneuvers, late reactions to a driving situation).

> *Hypothesis 3: Trust in an automated driving system will be influenced by the experience of events of low system reliability.*

Regarding the interaction with the two other main factors, it was assumed that experience and system transparency both help to prevent the decrease in trust when low reliability of the system is experienced (e.g., see Dzindolet et al., 2003).

- 3a) *Experiencing more than one event of low system reliability will lead to a decrease in trust in the system.*

3b) *Experience will help to understand system behavior and boundaries, and will thus reduce the decline in trust after events of low system reliability.*
3c) *Detailed information will help to understand system behavior and boundaries, and will thus reduce the decline in trust after events of low system reliability.*

The presented hypotheses were derived from outcomes of the first user study and were answered with the help of the second study. This study is presented in the following (Section 4.3.2).

4.3.2 Study design

To examine the hypotheses introduced above, the second user study focused on the variation of system performance to investigate the effect of system reliability on trust in automated vehicles in a structured analysis. In the context of HAD, the investigation of low system reliability has the potential of being hazardous, because situations including system boundaries need to be created. Even though drivers have a time reserve until they need to react to a takeover request, participants should be protected from any potential risks. Therefore, the second user study was set up as a simulator study. The simulated environment and the general approach adopted in the user study are described in this section.

Participants

As in the first user study, participants in the second study were recruited from employees of Volkswagen Aktiengesellschaft. They received a small gift as an incentive for volunteering as test persons. 35 women and 37 men took part in the simulator study, with an average age of 37.25 years ($SD = 10.17$ years) and an average mileage of 17 015 km of driving per year ($SD = 10337$ km). Twelve of the 72 participants preferred being driven as a passenger over driving themselves. On average, participants considered themselves slightly better than other drivers.

Simulator setting

It is considered difficult to measure a variable like trust, which is per definition arising only in situations of uncertainty and vulnerability (Lee & See, 2004, see Section 2.2.1), in a simulated environment, where at no time the driver needs to feel unsafe and where the immersion into the situation is limited. De Winter, Van Leeuwen, and Happee (2012) rightly criticize that in a driving simulator, no real consequences or dangers of an action can occur, thus provoking a false sense of safety. This limitation in the stimulation of relevant perceptual cues of the participants could lead to biased ratings of perceived trust, acceptance, or safety (Albert, Lange, Schmidt, Wimmer, & Bengler, 2015). However, to be able to investigate the effect of boundaries of an automated system in the context at hand, it was necessary to use a controlled environment to avoid compromising the participants' safety. One advantage of the simulated environment is the harmless presentation of situations that would potentially be critical in reality (De Winter et al., 2012). Also, in the simulator one has the capability of showing a specific traffic scenario repetitively and predictably (Albert et al., 2015). In addition, a setting in a controllable environment can of course reduce the impact of other confounding variables, supporting a structured analysis of selected factors.

The simulator software Virtual Test Drive (VTD) by VIRES Simulationstechnologie GmbH (2014) was used for the simulated environment in this study. With the help of this environment, visually realistic traffic situations with a surrounding landscape, a road network consisting of highway, rural and urban roads, and other vehicles can be recreated. All road users can be manipulated to react in a specific way, and otherwise will act self-determinedly following realistic motion patterns. The environment was presented on three projection screens of 3.0 x 2.3 meters size in the front and three 42-inch LCD monitors in the back to provide a fairly complete picture of the situation. In addition, driving noises were emulated to increase the realism of the scenery. All settings could be adjusted by the experimenter in the observer room that can be seen in Figure 4.20 (left).

Figure 4.20. Setting of the second user study in the simulator with observer room (left) and static driving simulator with projection screens (right).

The mock-up used to drive in the virtual world included a full cockpit resembling an Audi Q7 with two front seats and a center console, as Figure 4.20 (right) shows. The interaction concept resembled the concept introduced in Section 3.3.2 (see Figure 3.6). A steering wheel with two dedicated buttons for the automated driving system (5), customary pedals, and a freely programmable instrument cluster (1) were installed in the mock-up, as well as an LED bar below the windshield (3). The setting of the mock-up in the simulator room can be found in Figure 4.21.

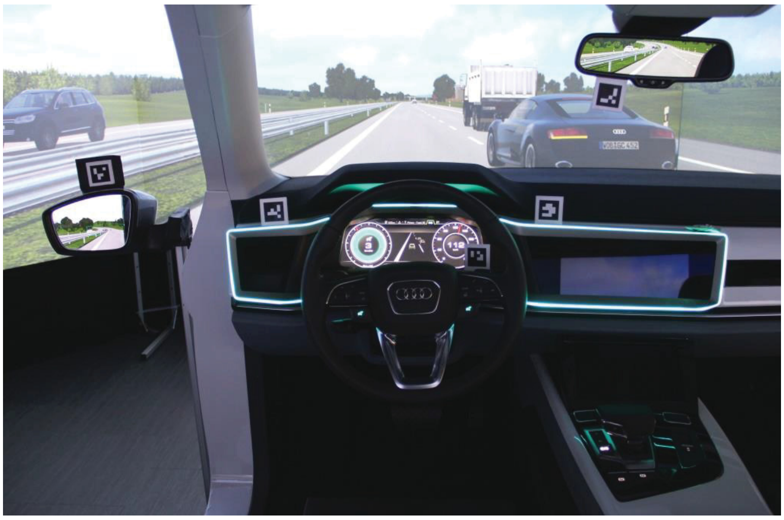


Figure 4.21. Setup of the mock-up for the simulated drive.

Study design and independent factors

In a 2 (experience with the system) x 3 (system transparency) x 2 (reliability) mixed factorial design, participants experienced the automated driving system in different simulated driving situations. Two independent between-subjects factors were varied. Firstly, it was differentiated between a group of participants experienced with the system and a group of novice users. As a second factor, the information given in an interface was varied (similar to the first user study presented in Section 4.2). In addition, reliability was varied as a within-subject factor. In five of the eight driving scenarios, the system mastered the driving situations well, but in three scenarios it did behave in an unexpected or unreliable way. The complete study design and procedure is visualized in Figure 4.25.

Level of experience of the user. The level of experience participants had with the HAD system was varied to see whether experience with a system leads to a different handling and understanding of the system, thus also altering the perception and appraisal of information given during a highly automated drive. Participants of the *experimental group* were introduced to the automated driving system in advance of the actual experimental drive. They were invited to experience the system on a 15 minute test drive, during which they received an extensive explanation of the system by the experimenter. The experimenter explained the system's capabilities and behavior, and how the displays can help to understand the system during the highly automated drive. Participants in the *control group* did not receive a special introduction, but only had a five minute test drive in the simulated environment without any explanations about the system.

System transparency. The importance of system transparency was investigated in more detail in this second user study, paying closer attention to events of low system reliability. Especially in situations where the system does not behave as expected by the driver, it was assumed that the HMI can help to avoid irritation and misunderstanding. Three different levels of transparency were distinguished (see Figure 4.22). As a control group, a status indication (A) was designed to only give information about the activation status of the automated driving system (active or passive) and how long the system is available (Figure 4.22, top). A more detailed HMI version (B) additionally showed the detected surroundings (Figure 4.22, middle). A comprehensive version of the HMI (C) displayed every sensory information and every behavior of the system, including maneuvers and reasons for a maneuver (Figure 4.22, bottom).

Reliability of the system. In the simulated environment, it could better be controlled for the occurrence and kind of events of low system reliability. Participants all experienced eight different driving situations while in HAD mode. Five of these runs included traffic situations which were well solved by the HAD vehicle (e.g., merging car directly in front of the ego vehicle, accident on the right lane). These high reliability scenarios contained the traffic situations depicted in Figure 4.23.

The three remaining scenarios included difficult traffic situations that the system did not solve ideally and where unanticipated system reactions occurred. None of these unreliable driving scenarios required the participant to take control in order to avoid a crash. The system would always work flawless again after a couple of seconds. One of the unreliable driving scenarios constituted a *miss*, where the HAD vehicle started an overtaking maneuver and canceled it in the last second because of upcoming traffic from behind (Figure 4.24, scenario 6). The vehicle returned to the previous lane without intervention from the driver. The second situation was a *false alarm*, where the automated driving system mistakenly detected a slow vehicle in front and performed an unnecessary overtaking maneuver (Figure 4.24, scenario 7). In the third situation of low reliability, the system misidentified the hard shoulder as a lane and started driving on the hard shoulder for approximately 5 seconds before returning to the actual lane (Figure 4.24, scenario 8). This situation may be interpreted as an external error that might happen in reality when lane markings are missing.

The unreliable driving scenarios were designed to represent system boundaries of the real system. Both impossible and unnecessary overtaking maneuvers could happen with the prototype

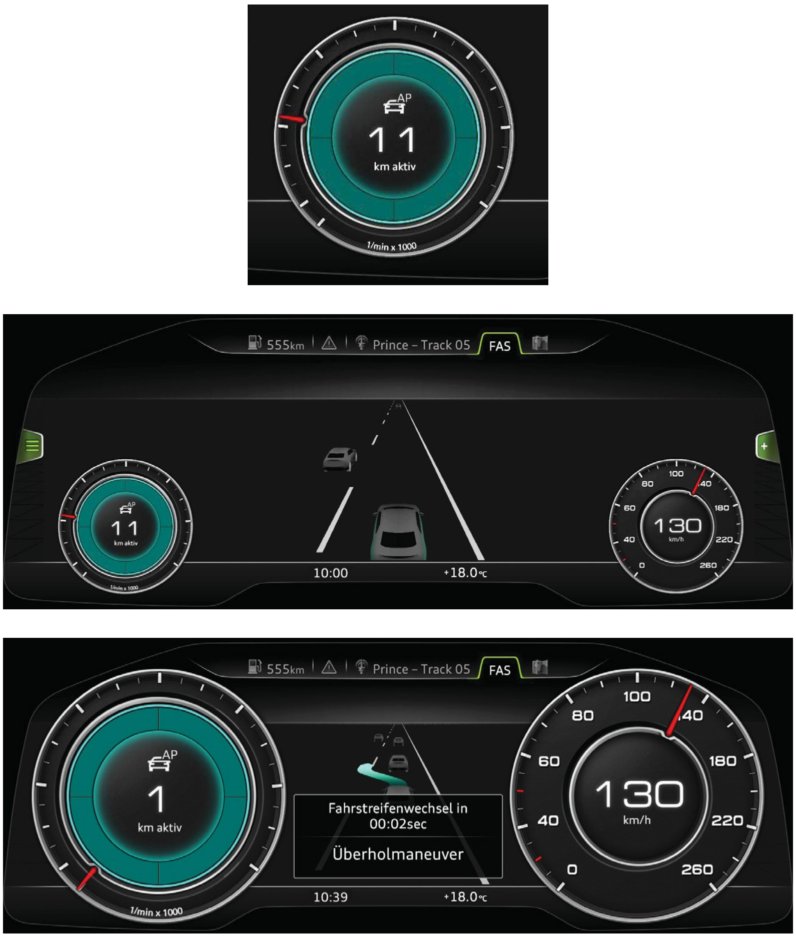


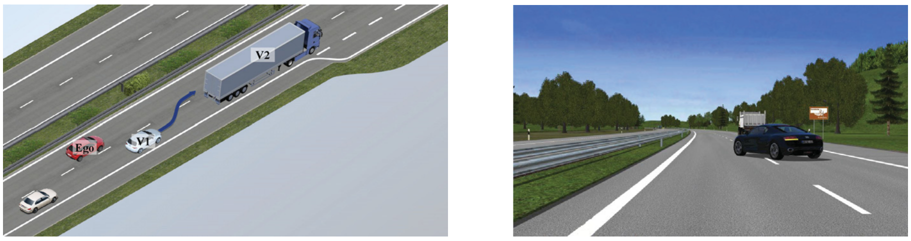
Figure 4.22. HMI versions A (top), B (middle), and C (bottom) used in the second user study.

vehicle, as well as difficulties with finding the correct lane in case of missing lane markings. The type of reliability of the first drive was controlled, to avoid first-experience effects. Half of the participants first experienced a reliable system, and the other half first rated a run with low system reliability. The rest of the scenarios were driven in randomized order.

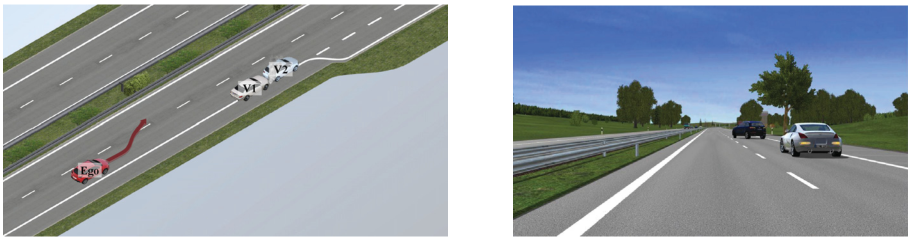
Dependent and mediating variables

As in the first user study, subjective and objective data were collected during the study. To collect subjective assessments of the participants with questionnaires, a windows tablet was utilized.

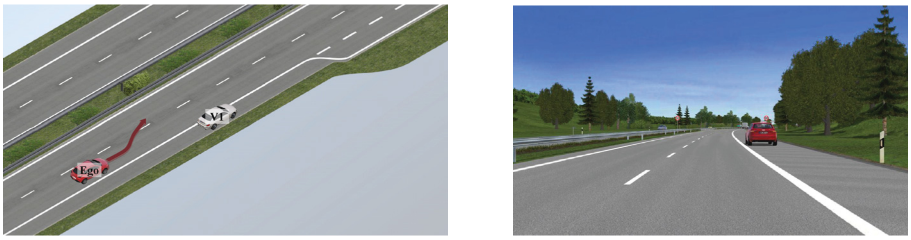
Subjective trust. Again, the questionnaire of Lee and Moray (1992) (adapted version from Muir, 1989) was used for trust assessments, including the four items addressing predictability, dependability, faith, and overall trust. The questionnaire can be found in Appendix A.1.3,



Scenario 1: Merging vehicle in front of own vehicle

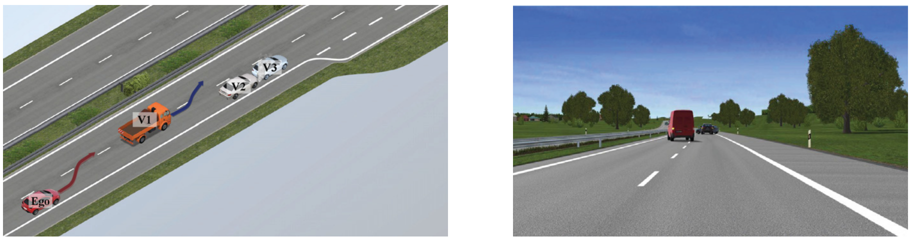


Scenario 2: Accident site on the right lane



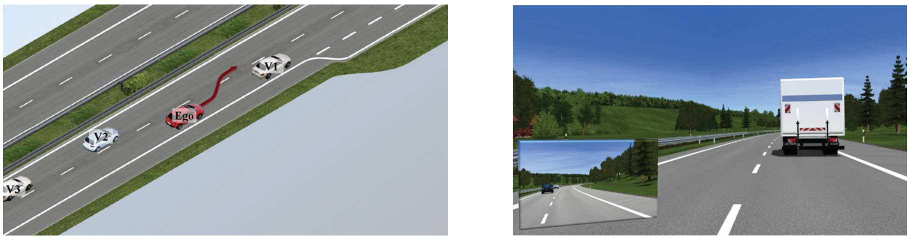
Scenario 3: Broken-down vehicle on the hard shoulder

Scenario 4: Highway traffic jam

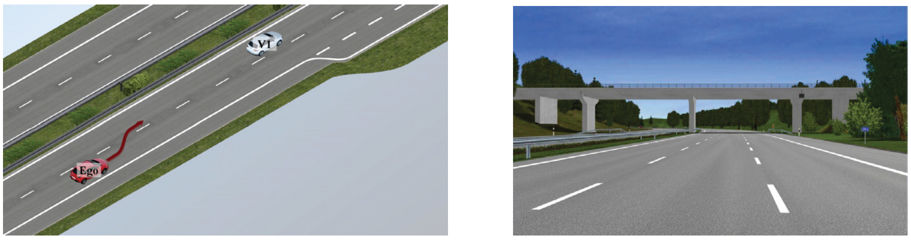


Scenario 5: Collision of two vehicles, blocked from sight

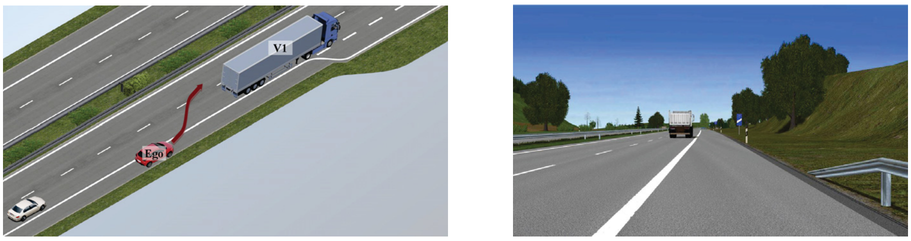
Figure 4.23. Reliable driving scenarios of the second user study.



Scenario 6: Failed overtaking maneuver



Scenario 7: Unnecessary overtaking maneuver



Scenario 8: Driving on hard shoulder

Figure 4.24. Unreliable driving scenarios of the second user study.

Table A.5. The items were answered by participants after each of the eight scenarios. A 15-point rating scale (from 1 = 'very low' to 15 = 'very high'; Heller, 1985) was utilized (see Appendix A.1.2, Table A.2).

Evaluation of HMI. Usefulness, relevance, and understandability of the HMI were evaluated after the training drive (also on the 15-point scale). Usefulness of the information was also rated after each drive to reveal changes over time.

Evaluation of the test drives. Perceived performance of the system and nervousness during the drive were also rated after each drive. Also here, the 15-point scale was used (see Appendix A.1.3, Table A.8).

Driving parameters. Relevant driving parameters were recorded during the simulated drive. Parameters of interest were related to the driver's interaction with the vehicle (e.g., gas and brake pedal position, steering wheel angle and hands-on detection) and to the system's states (e.g., current state, executed maneuvers). An analysis of these data enabled a calculation of handover and takeover times of the participants to indications of the automated system.

Gaze behavior. To see where participants focus their attention during a highly automated drive and if a special interest arises in certain information during low system reliability, behavioral data collection included the participants' focus of attention. The attention allocation of the drivers during HAD was investigated with data about the participants' gaze behavior. Again, the head-mounted Dikablis eye-tracking system (Ergoneers GmbH, 2015, see Section 4.1.4) was utilized for data collection. The relevant AOIs were the street, side and rear view mirrors, and the instrument cluster with information on the automated driving system (see Figure 4.2).

Procedure

Participants were invited to take part in a two-hour simulator study. In the beginning, participants were invited to take a seat in the mock-up and the experimenter briefly explained the planned course of the study. Subsequently, participants were asked to answer initial questions on demographic and individual characteristics on the tablet. After they had finished the questionnaire, participants put on the eye tracking system for the collection of gaze data and the first training drives were started. The instructions for the second user study can be found in Appendix A.1.1. As already mentioned, the level of experience with the system was varied by differing introductions to the system. The control group carried out a five-minute training drive, and was only introduced to the basic interaction with the HAD system in form of the dedicated buttons on the steering wheel to engage or disengage the system. The experimental group received a more comprehensive explanation about the HAD system during a fifteen-minute training session, including the functionality and capabilities of the system and the interfaces displaying specific information about the automated drive. After the training drives were concluded, participant gave an assessment of their first impression of the interface concept regarding usefulness, relevance, and understandability. Then, the actual test scenarios were started.

Each test scenario consisted of a highway service station, where participants started to drive, and a round course with several kilometers of highway road. Each drive had a duration of approximately 5 minutes. Participants started in manual driving mode and were prompted to activate the HAD system with the two buttons on the steering wheel as soon as they were on the highway road. As under real traffic conditions, the automated driving system could control speed, distance to other vehicles, and lane change maneuvers. Thus, the participants could release all controls of the vehicle and did not need to take any action until the system indicated the proximity of the end of the automated driving stretch of the road by a visual indication together with an acoustic announcement and sound. For an immediate deactivation of the system, participants only needed to press the two dedicated buttons, the brake, or turn the steering wheel. The test scenarios were presented in randomized order. In five of the eight scenarios participants encountered, the automated driving system reacted reliably to every situation throughout the drive (high reliability scenarios). In the remaining three scenarios, the system did not react reliably to the driving situations (low reliability scenarios). After each run, participants rated their perception of the system's performance, their nervousness during the drive and their trust in the automated driving system. After the last scenario was completed successfully, participants filled out a final questionnaire and were thanked for their participation. The whole course of the second user study can be found in Figure 4.25.

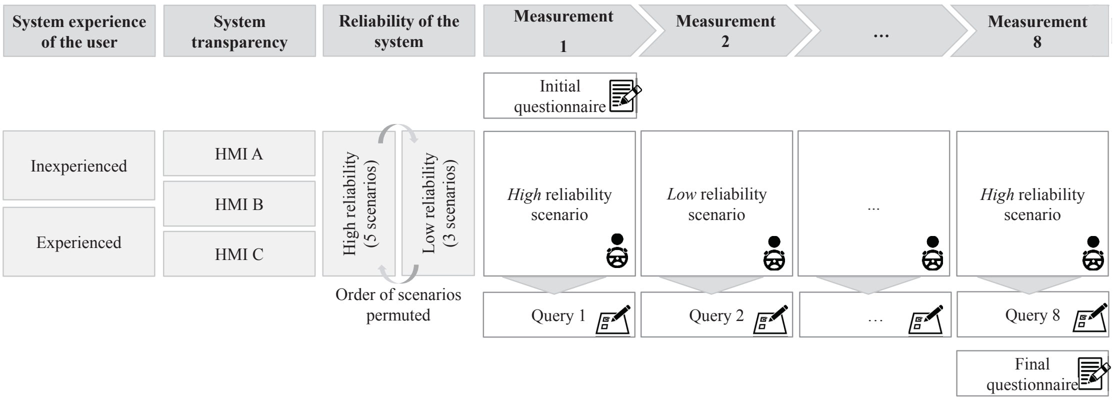


Figure 4.25. Procedure of the second user study.

4.3.3 Results

The statistical analysis of the data was again done using IBM SPSS Statistics (version 19.0), and statistical figures were created with R (version 3.4.1). The independent variables were analyzed statistically regarding the postulated hypotheses. Considering trust, again the four-item questionnaire of Muir (1989) was used for each drive and an overall ranking was computed out of these queries. The dependent variables were checked for normal distribution with the Shapiro-Wilk test, and extreme outliers were removed from the analyses. The main factors as well as covariates were analyzed in detail with independent t-tests or analyses of variance. An alpha level of $\alpha = .05$ was defined as significant for the analyses. Table A.15 in Appendix A.2.1 gives an overview over the correlations of the main variables.

Hypothesis 1: Experience with the system

The first hypothesis assumed that prior knowledge about or experience with an automated driving system can increase the level of trust in the system. A group of participants who was trained to use the system (hereafter referred to as experienced users) was compared to a group which used the system without further information (referred to as inexperienced users). The second hypothesis postulated an effect of system transparency. The two between-subjects factors were analyzed in a 2 (experience) x 3 (system transparency) test design with an analysis of variance.

Trust. As a first step, the overall development of trust was examined. No rise in trust was in evidence, although the system performed well in the majority of the drives. On the contrary, a drop in trust was detected after the first drive, $F(7,497) = 3.06$, $p = .004$, $\eta_p^2 = 0.04$, compared to all other drives (paired comparison $p_{1,2} = .011$, $p_{1,3} = .020$, $p_{1,4} < .001$, $p_{1,5} = .039$, $p_{1,6} = .027$, $p_{1,7} = .052$, $p_{1,8} = .005$). Figure 4.26 shows the development of trust over the course of the eight drives.

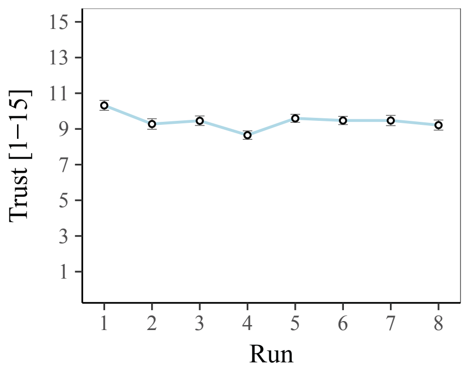


Figure 4.26. Mean trust ratings during HAD. Error bars represent standard errors.

A higher level of trust was predicted for people with experience with the automated driving system. Trust ratings of the two experience groups were compared, analyzing the ratings of the first drive as well as all drives taken together (see Figure 4.27). With the two-factor analysis of variance, no differences in the reported level of trust could be found between inexperienced and experienced participants (first drive: $F(1,66) = 0.05$, $p = .826$, all drives: $F(1,66) = 0.39$, $p = .536$), contrary to the assumption.

Nervousness. The expected gain of trust due to experience with the automated driving system was assumed to go along with a decline in the level of stress or nervousness over the course of the study. The level of stress after the first drive and over all drives was compared between the two experience groups (Figure 4.28). Surprisingly, a significant difference opposing the assumption could be found for initial nervousness ratings, $F(1,66) = 10.64$, $p = .002$, $\eta_p^2 = 0.14$. Drivers with experience with the automated driving system were more nervous in the beginning of the test drive. No effect was found over all drives, $F(1,66) = 1.89$, $p = .174$.

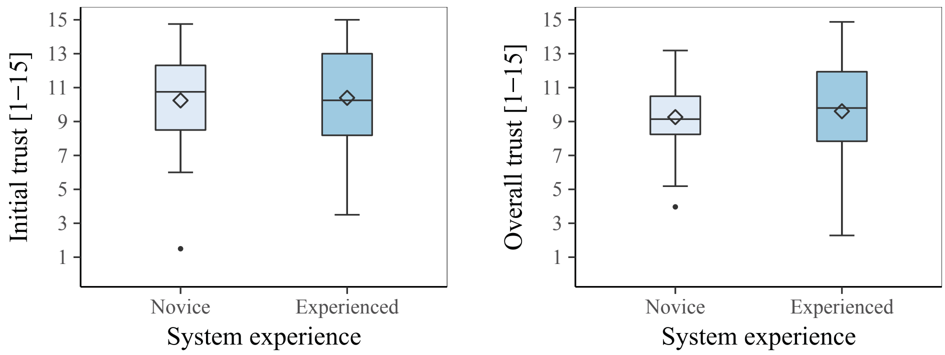


Figure 4.27. Initial (left) and overall (right) trust ratings of the two experience groups for the HAD system. The boxplots show the trust ratings of the inexperienced and experienced user group. Error bars represent the standard deviation.

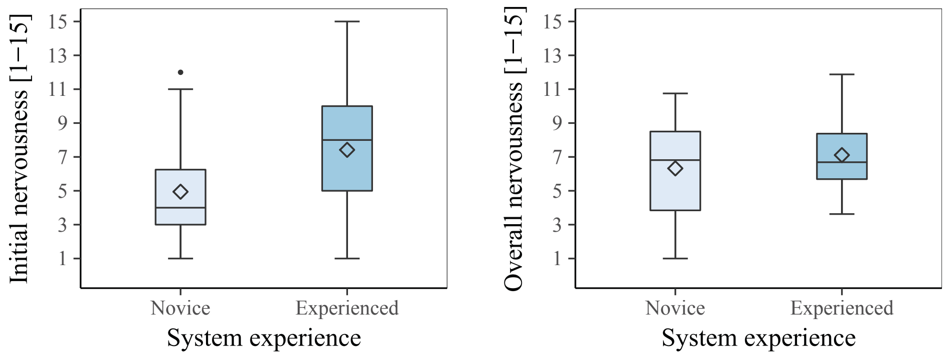


Figure 4.28. Initial (left) and overall (right) nervousness of the two experience groups while driving with the HAD system. The boxplot shows the nervousness of the inexperienced and experienced user group. Error bars represent the standard deviation.

Evaluation of HMI. Furthermore, it was suggested in sub-hypothesis 1c that a higher level of experience with the automated driving system will decrease the perceived usefulness of the interface over time. However, this effect could not be found in the data. On the contrary, results point in the direction of a higher perceived usefulness rating of the experienced participants for the interface concept after the first drive, $F(1,66) = 4.84$, $p = .031$, $\eta_p^2 = 0.07$, and over all drives, $F(1,66) = 3.87$, $p = .053$, $\eta_p^2 = 0.06$, compared to the inexperienced group (Figure 4.29).

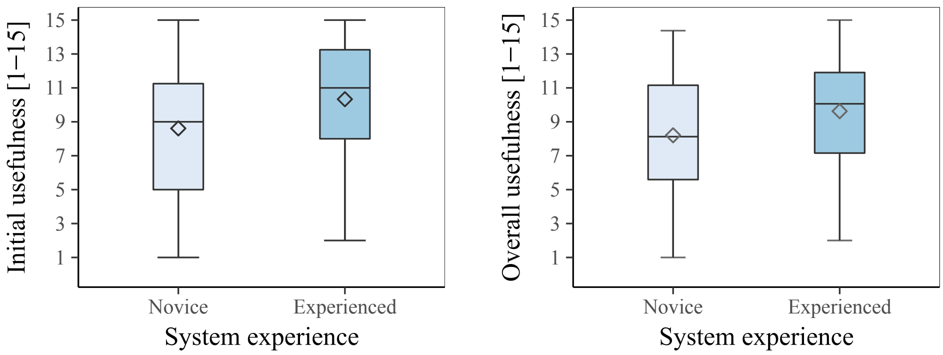


Figure 4.29. Initial (left) and overall (right) usefulness ratings for the interface concept of the two experience groups. The boxplot shows the usefulness ratings of the inexperienced and experienced user group. Error bars represent the standard deviation.

Hypothesis 2: System transparency

In the second hypothesis of the user study, it was assumed that system transparency, realized by an HMI providing information about the system's state and behavior to the driver, can influence trust in the automated driving system when the performance of the system is not as desired. With the second part of the two-factor analysis of variance this hypothesis was analyzed in detail.

Trust. To see if trust is influenced by the transparency of the system, initial and overall trust ratings for the three different HMI versions were compared. With analyses of variance, no statistically significant differences could be identified in initial or in overall trust ratings (first drive: $F(2,66) = 1.09$, $p = .343$, over all drives: $F(2,66) = 0.85$, $p = .431$), as Figure 4.30 shows on a descriptive level. Also, no interaction effect was found between the two factors experience and system transparency.

Nervousness. Furthermore, it was assumed that the level of stress will decline when more information about the system is available, thus when the system has a higher transparency. The nervousness ratings after the first drive and over all drives were compared depending on the three HMI versions. Results did not show a significantly higher stress level when less information was presented, $F(2,66) = 0.62$, $p = .540$ (first drive), $F(2,66) = 0.22$, $p = .807$ (over all drives) (Figure 4.31). Again, no interaction effect was found.

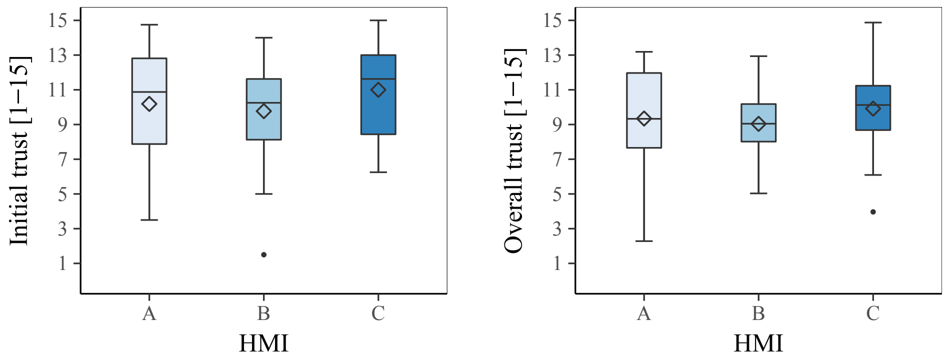


Figure 4.30. Initial (left) and overall (right) trust ratings for the three different interface concepts. The boxplot shows the trust ratings for the three different interfaces A, B, and C of the HAD system. Error bars represent the standard deviation.

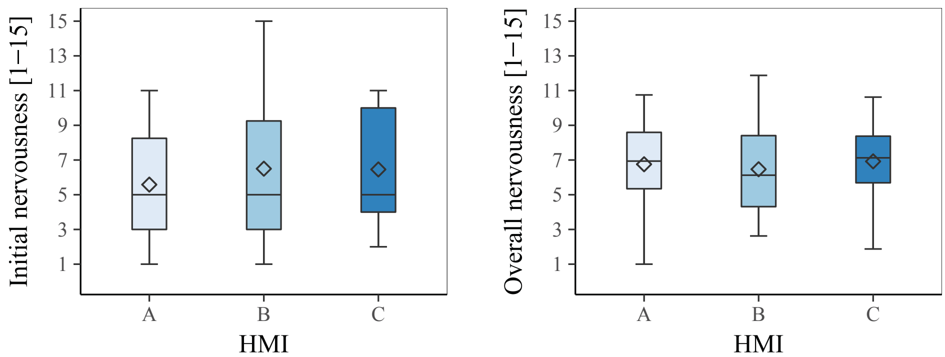


Figure 4.31. Initial (left) and overall (right) nervousness ratings for the three different interface concepts. The boxplot shows the results of the nervousness ratings for the three different interfaces A, B, and C. Error bars represent the standard deviation.

Evaluation of HMI. It was expected that even though drivers were driven automatically, a high need for information would result in a higher usefulness rating for the more detailed information given in HMI C. Indeed, HMI A received significantly lower ratings for usefulness after the first drive, $F(2,66) = 13.93$, $p < .001$, $\eta_p^2 = 0.30$, as well as over all drives, $F(2,66) = 11.11$, $p < .001$, $\eta_p^2 = 0.25$, compared to the other two HMI versions. This effect was revealed with paired comparisons (initial usefulness: $p_{A,B} = .001$, $p_{A,C} < .001$; overall usefulness: $p_{A,B} = .001$, $p_{A,C} < .001$), which are visualized in Figure 4.32.

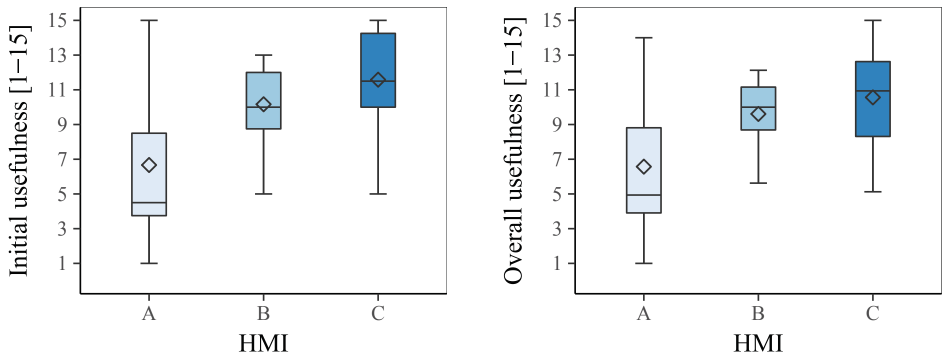


Figure 4.32. Initial (left) and overall (right) usefulness ratings for the three different interface concepts. The boxplot shows the results of the usefulness ratings for the three different interfaces A, B, and C of the HAD system. Error bars represent the standard deviation.

To conclude, the two main factors and their interaction did not reveal any significant differences in trust. Neither the main effects of experience with the system or system transparency nor an interaction effect were statistically significant (Figure 4.33).

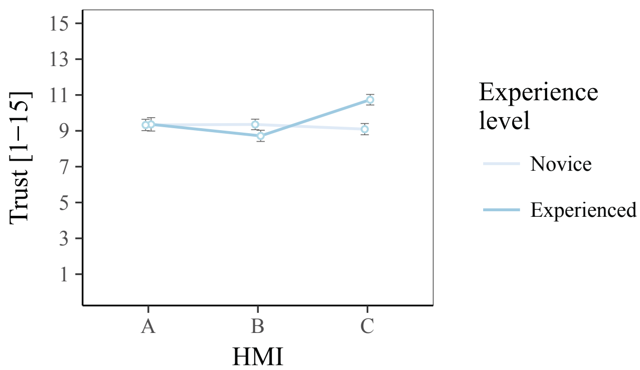


Figure 4.33. Mean trust ratings for the three different interface concepts depending on the user's level of experience. The line graph shows the results of the trust ratings for the three different interfaces A, B, and C of the inexperienced (light blue) and experienced (dark blue) user group with the HAD system. Error bars represent standard errors.

As an addition to the questionnaires, participants were also asked to choose their preferred concept by ranking the HMI versions in an order of priority. With nearly 70%, HMI C was preferred by the majority of participants, $\chi^2(2, N = 72) = 45.08$, $p < .001$ (see Figure 4.34).

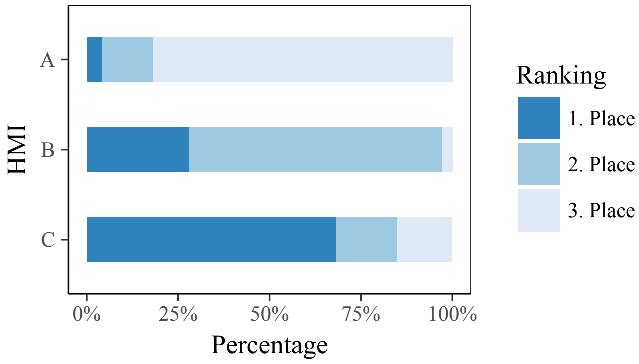


Figure 4.34. Results for subjective HMI preference in the second user study. The bar plot shows the percentage distribution of ratings for the three HMI concepts.

Hypothesis 3: System reliability

The third hypothesis dealt with the influence of system interaction and the effect of low system performance on trust in the automated driving system.

System reliability and trust development. In the first sub-hypothesis 3a, it was postulated that several events of low system reliability will decrease trust in the automated driving system. To verify this assumption, the three scenarios containing an event of low system reliability were analyzed. The trust ratings stemming from the first, second, and third unreliable drive were compared to the overall trust rating. The repeated measures analysis showed a significant main effect of the factor drive, $F(2.13, 151.06) = 10.27$, $p < .001$, $\eta_p^2 = 0.13$ (corrected with Greenhouse-Geisser), and the paired comparisons revealed that the significant difference was between the third unreliable drive and the overall trust rating ($p_{3,overall} < .001$). As expected, trust ratings for the preceding drives did not differ significantly from the overall trust rating.

Interestingly, an exploratory inspection of the performance ratings for all scenarios also revealed an additional effect. In scenarios 1 to 5, the vehicle performed highly reliable, while in scenarios 6 to 8, it behaved in an unexpected way. In Figure 4.35, performance ratings of all scenarios are compared. A repeated measures analysis of variance revealed specific differences between the scenarios, $F(5.78, 410.53) = 28.88$, $p < .001$, $\eta_p^2 = 0.29$ (Greenhouse-Geisser corrected). Of the scenarios 6, 7, and 8 with low system reliability, only in the scenarios 6 and 8 the system's performance was actually perceived as low. Perceived performance in scenario 7 did not differ significantly from most of the reliable scenarios, as simple contrasts showed ($p_{7,1} = .002$, $p_{7,2} = .567$, $p_{7,3} = .196$, $p_{7,4} = .445$, $p_{7,5} = .484$, $p_{7,6} < .001$, $p_{7,8} < .001$). In contrast to scenarios 6 and 8, scenario 7 included a false detection (and an unnecessary ma-

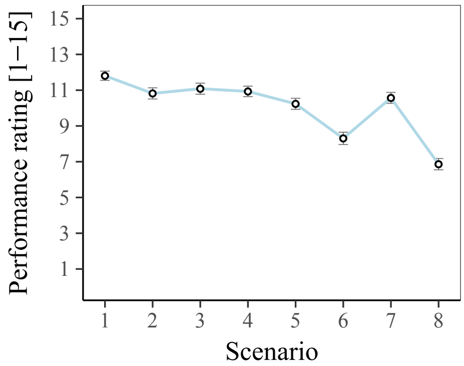


Figure 4.35. Mean performance ratings for all scenarios. In scenarios 1 to 5, the vehicle performed highly reliable, while in scenarios 6 to 8, it behaved in an undesired way. Error bars represent standard errors.

neuver) of the system, not a miss of the detection. This scenario seems to have been perceived differently and system performance was thus not rated low. Consequently, participants actually experienced only two drives with a low reliability of the system. The other scenario with low reliability of the system was not perceived as a bad performance of the system.

System reliability, experience, and system transparency. To investigate how the two other main factors (experience and system transparency) interact with the factor system reliability, the whole model consisting of three factors was analyzed. A repeated measures analysis of variance for trust ratings in the reliable and unreliable scenarios revealed a statistically significant main effect of the repeated measures factor reliability, $F(1,66) = 70.45$, $p < .001$, $\eta_p^2 = 0.52$. As analyzed before, the between subjects factors experience and system transparency did not show significant differences, and no interaction effect between system reliability and transparency could be found. An interaction effect between system reliability and experience was found to be significant, $F(1,66) = 5.54$, $p = .022$, $\eta_p^2 = 0.08$. Thus, when reliability of the system was low, trust of the experienced users was not reduced as much as the trust of the inexperienced group (see Figure 4.36).

Behavioral parameters

As an additional, exploratory analysis, driver behavior was again taken into account. To re-examine if also in the second user study, reliant behavior can be found as a consequence of trust in the automated system, percentage of gazes to different AOIs (street, instrument cluster, and mirrors) was analyzed to interpret attention allocation. As in the first user study, percentage gaze

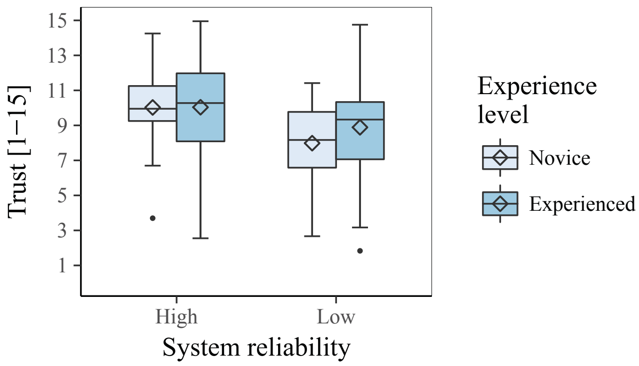


Figure 4.36. Trust of the two experience groups depending on system reliability. The boxplot shows the results of the trust ratings of the inexperienced (light blue) and experienced (dark blue) user group depending on low or high system reliability. Error bars represent the standard deviation.

distribution (attention ratio) was gathered over all runs. When comparing the attention allocation of inexperienced and experienced users, an analysis of variance revealed a significant difference in gazes on the instrument cluster, *Welch's* $F(1,57) = 5.77$, $p = .020$, $\eta_p^2 = 0.78$. Inexperienced users put more attention on the instrument cluster than experienced users. Also, a significant difference in gazes on the instrument cluster was found with an analysis of variance between the three HMI concepts, $F(2,65) = 4.66$, $p = .013$, $\eta_p^2 = 0.13$. Paired comparisons showed significant differences of HMI A compared to HMI B and HMI C, with less percentage of gazes on the instrument cluster ($p_{A,B} = .005$; $p_{A,C} = .021$). Lastly, as Figure 4.37 shows, a comparison between pooled trust groups (median split) indicated a tendency that participants trusting the system less located more attention on the instrument cluster than participants with higher trust in the system, $F(1,66) = 3.33$, $p = .073$, $\eta_p^2 = 0.48$. Bigger circles indicate more attention than smaller circles on an AOI. The orange box and the asterisk mark the significant result.

Another behavioral parameter that is related to attention allocation is the reaction to takeover requests during automated driving. Someone whose attention is not focused on the driving situation and the behavior of the automated driving system is assumed to need more time to take back control from the vehicle. Drivers not trusting the system are assumed to observe the vehicle's behavior carefully and intervene whenever they feel unsafe. No significant differences were found for the main factors experience and transparency regarding the hands-on reaction time (time until participants touched the steering wheel) and the takeover reaction time (time until deactivation of the system) after a takeover request. Also, no difference in reaction times was found between the group of participants trusting the system less and the group trusting the system more. Results

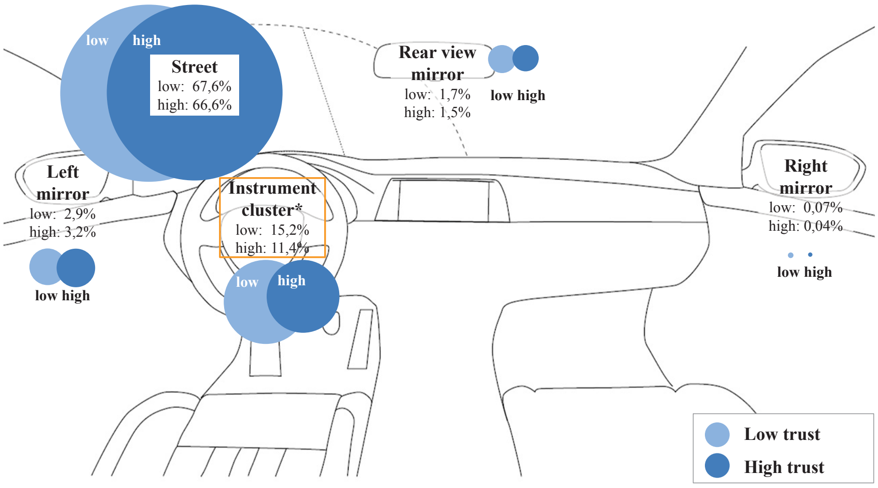


Figure 4.37. Percentage of gazes (attention ratio) depending on low trust (light blue) or high trust (dark blue) in the automated driving system. The size of the circles is relative to the percentage of gazes in this area.

regarding the number of takeovers initiated by the participants, by contrast, revealed (marginal) significant differences for the main factor experience, $t(70) = 1.77$, $p = .081$, $d = 0.42$, as well as for the two trust groups, $t(70) = 3.06$, $p = .003$, $d = 0.72$. Participants with experience with the automated driving system took back control more often ($M = 2.64$, $SD = 3.01$) than the control group without experience ($M = 1.56$, $SD = 2.12$) (Figure 4.38, left). Furthermore, participants of the group with high trust ratings did not interrupt the automated driving system as often ($M = 1.17$, $SD = 1.89$) as the group with lower trust ratings ($M = 2.97$, $SD = 2.96$) (Figure 4.38, right).

4.3.4 Implications

The second user study was designed to investigate how trust is influenced by unanticipated system reactions and whether it can be supported and stabilized through system transparency with the help of an HMI concept. To be able to investigate events of low system reliability in an automated vehicle, this study was conducted in the safe environment of a driving simulator. 72 participants evaluated the system's transparency and reliability during eight different traffic scenarios. The study focused on the examination of individual characteristics of the driver (*experience*) in conjunction with the system characteristics *reliability* and *transparency*.

Findings. The first hypothesis postulated an influence of system experience on trust in an automated driving system. System experience was realized through a training session with de-

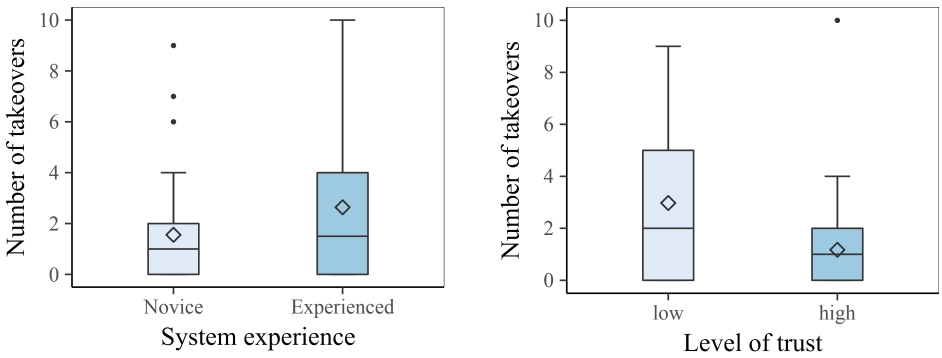


Figure 4.38. Number of takeovers depending on the level of experience (left) and the level of trust in the system (right). The boxplot shows the number of takeovers depending on a low vs. high trust or experience. Error bars represent the standard deviation.

tailed explanations about the automated driving system and the interface. Participants with more experience with the system did not develop more trust in the system in the beginning of the interaction or during the course of the study. Even more, experienced participants stated to be more nervous than the control group during the first drive with the system. Hypotheses 1a and b are not supported by the data. Results showed no effect of experience on trust ratings, and experience did not result in less stress for the drivers. It can be speculated that experience led to a better understanding not only of the capabilities of the system, but also of its boundaries (indicated through takeover requests of the system). Possibly, inexperienced drivers rather assumed that the automated driving system will manage every situation flawlessly. This attitude towards the automated system can be attributed to a positive automation bias, a concept formerly suggested by Mosier and Skitka (1996), Mosier et al. (1998), as well as Dzindolet et al. (2003). It suggests that inexperienced system users will expect an automated system to perform reliably, leading to a high initial trust in the system. Trust is then calibrated during system use and thus declines after the first experience with the system. This conclusion is furthermore supported by the observation that a decline in trust due to low system reliability was more serious for inexperienced participants. Experienced participant found the interface concept more useful than the control group, indicating that with more interaction with the concept, information about the automated system becomes more useful and important. Hypothesis 1c had assumed that information gets less relevant with more system experience. No proof of this assumption was found. Experienced users had the same need for information as novice users and rated the HMI just as useful (especially the more comprehensive versions B and C). One possible explanation for this result is the very short familiarization phase with the system: even the users in the experimental group did only try the system for 15 minutes before the test runs started. This level of experience might not have been high enough to reduce the need for information, although experienced drivers did understand the indications of the HMI better and rated them as more relevant. Overall, the first hypothesis needs to be rejected.

In the second hypothesis, a positive influence of transparent system indications on trust development was postulated. As in the first user study, no general effect of transparency on trust in the automated driving system could be verified. Also, the level of nervousness was not influenced by the transparency of the system. However, subjective ratings regarding the usefulness of the concepts indicated a clear preference for the most comprehensive HMI C. Also when ranking the concepts, the majority of participants were in favor of HMI C. It can thus be concluded that drivers want to get detailed information about the drive and the vehicle's planned behavior—although this information does not seem to influence their trust in the automated vehicle. Thus, only hypothesis 2c can be accepted.

The other main factor that had been identified in the first study, namely perceived system performance, was varied in the second study to enable a structured analysis. The assumption in the third hypothesis stated that trust in an automated driving system will rise with growing experience with the system, but that experiencing several unanticipated system reactions will lower trust in the automated system. In the course of the study, participants experienced several different scenarios, partly with unanticipated system behavior, that lead to an adjustment of trust ratings. Interestingly, however, participants distinguished between different levels of low reliability of the system: 'misses', that could potentially have a harmful outcome (though they did not have in the simulated scenario), were rated worse in subjective performance ratings, while a 'false alarms' or rather a false behavior with only an unnecessary, but not hazardous action was rated nearly as good as the scenarios with high system performance. The mere fact that the system erred did not automatically lead to a loss in perceived performance. This result goes along with the findings of Dixon et al. (2007), who showed that misses and false alarms have a different effect on reliance. While these researchers found false alarms to be worse to overall performance during human-machine interaction, Masalonis et al. (1998) showed that a miss with potential consequences can lead to a more profound loss of trust than false alarms with less consequences. The results discussed so far prove that drivers are very well able to adjust their trust depending on their perceived performance of the system. Consequently, they are able to create an adequate (lower) level of trust depending on the system's capabilities, as suggested by hypothesis 3a. An exception seems to be the first experience of a faulty automation—in that case, drivers are forgiving. The experience of one of the two perceived low reliability events did not diminish trust significantly. As hypothesis 3b supposed, experience with the system can support trust when system boundaries are met. Unlike experience with the system, a higher transparency of the system did not help to dilute the effect of low system reliability, hence hypothesis 3c could not be confirmed. In the end, perceived system performance again was found to be the one factor influencing trust in an automated driving system the most, with experience being able to support trust in specific situations.

Limitations. When interpreting the results of the second user study, some reservations need to be made. As the second user study took place in a static simulator, it needs to be kept in mind that the immersion into the driving situation is not as deep as in a real driving environment. No actual risk of accidents was present, which may have had an effect on the participants' behavior. While the surroundings can be controlled relatively well and many confounding variables can be ruled out as influencing factors, the external validity is compromised due to the setting of the study in a simulated environment. Especially the feeling of driving automated is difficult to convey in a simulator, and trust is of less relevance in this environment. Because of this restriction, the reported trust in the automated driving system may not be fully comparable to results obtained in other environments, e.g., real driving settings. Also, the interior of the mock-up for HAD is not completely comparable to the interior the prototype vehicle, although the interaction concept was designed in the same way. Nonetheless, the reliability factor that was to be examined made the simulated environment necessary.

No research results were known regarding the time frame necessary to initially familiarize drivers with an automated driving system. The lack of difference in trust between the two experience groups might have been caused by a too short familiarization phase for the experienced group. A few more minutes of driving experience and the explanations of the experimenter seemed insufficient to create more trust in the beginning of the test drive. Only later in the course of the study, an effect could be found for situations of low system reliability. Also, while participants got to drive with the automated driving system for approximately one hour in total, driving sections were always interrupted by short stops to evaluate the system, hindering the normal flow of experience. However, the short breaks were necessary to be able to assess trust and other variables distinctly for each traffic scenario.

The less restricted simulator environment enabled a larger number of participants to take part in the study compared to the real driving studies. Participants were employees of the Volkswagen Aktiengesellschaft and might thus not be fully representative of a normal user group. However, they were not involved in the development of automated driving or HMI concepts to be as comparable as possible.

Conclusion. To sum up the findings of the second user study, it can be noted that in order to trust an automated vehicle, drivers need to be convinced of the automated vehicle's driving performance. Subsequently, the concrete level of experience with the automated driving system at hand is of relevance. For example, experience can be of help to classify events of low system reliability and to better put them into perspective. The assumed influence of the system's transparency on trust in the automated driving system was not supported by the data. However, the positive subjective ratings of the HMI concepts prevent an abandonment of the idea. Instead, thought needs to be given to the improvement of the design based on the evaluation and feedback of the users. It might furthermore be of interest to study the relevance of system transparency over several practical experiences with the system. This aspect has not been addressed in the user studies so far. First attempts were made when varying experience through initial explanations regarding the system. A next step was thus the actual investigation of prolonged system use and its impact on trust in an automated driving system.

4.4 Study 3: Beyond initial trust—Trust across multiple practical experiences

The third user study on trust in HAD aimed at the evaluation of an automated driving system during and after a longer time of usage, to broaden the knowledge already gained through the results of the first two user studies. In this user study, insights were not limited to the initial level of trust arising when interacting with the automated driving system for the first time. Long-term system trust was assessed over the course of four appointments, each with a trial duration of approximately two hours of driving in HAD mode. With up to eight hours of experience with the automated driving system in total, the objective of the user study was to reach a more complete familiarization with the system to eliminate the influence of novelty on the results. A longer time-frame was not possible due to limited access to the prototype vehicle. Trust development was gathered over the course of each appointment, with a query before, during, and after each trial, to depict the short-term as well as the long-term development of trust. The level of transparency of the automated driving system was part of the analysis again, as longer-term impacts of an HMI concept have (to the author's knowledge) never been investigated in the context of automated driving. Transparency was not found to be of main importance for initial contact with the automated driving system, as the vehicle's driving performance was more decisive. However, over the course of time and with more experience with the system, the system's HMI and its transparency might gain relevance.

4.4.1 Hypotheses

In the previous user studies no general positive effect of experience with the system on trust was apparent. This could be an effect due to the relatively short duration of system use in both preceding user studies. While short-term use of automated driving might not lead to an increase in trust in the system, mid- or long-term use can be expected to encourage trust in the system.

> *Hypothesis 1: Prolonged use of an automated driving system and gained positive experience with it increase the level of trust in the system.*

As the results of the previous user studies indicated, it was assumed that system trust would not rise after initial system use, and possibly even decline slightly due to a positive automation bias before the first use, as suggested by the results of the second user study. Trust was expected to rise with prolonged system interaction during the following trials.

> *1a) Initial use of an automated driving system will lead to a decline in system trust.*
>
> *1b) During later interaction with the automated driving system, trust will rise with increasing positive experience with it.*

An essential prerequisite for the expected increase in trust over time will be the adequate reliable performance of the automated driving system, as can be summarized from the first two user studies.

So far, the conducted user studies did not find significant influences of system transparency on trust in an automated driving system, although findings of other areas of research suggested an impact of this factor. One reason for the non-significant results could be the limited duration of system use. When interacting with an automated driving system for the first time, system performance and reliability might be of main importance. Interface design might get into the focus of attention with extended use of the system.

> *Hypothesis 2: High system transparency (information about the system's state and behavior) supports the growth of trust during prolonged system use.*

It was suggested by the previous results that system transparency does not have an influence in the beginning of system use, as other factors (like system performance) are far more important for initial contact with the system. Transparency was thus not expected to be able to increase trust in the automated driving system during the first drives. However, after prolonged system use the influence of performance was expected to decrease, giving way for transparency to gain influence and support system trust.

2a) *Detailed information about the system's behavior will not be relevant for trust during initial interaction with the system.*

2b) *Detailed information about the system's behavior will help to create trust during prolonged interaction with the system compared to less detailed status information.*

Again, the assumption was made that a rising level of trust will become apparent in the form of observable behavior.

Hypothesis 3: Trust in an automated driving system goes along with a shift of attention away from the driving task.

Specifically, it was expected that the attention focus of the driver and their likelihood to engage in non-driving-related activities will change depending on the duration of system use, the HMI concept, and the level of trust the driver has developed in the system.

3a) *Over time, a person's attention diverts away from the driving situation.*

3b) *With more detailed information about the system's behavior, a person's attention diverts away from the driving situation.*

3c) *With higher trust in the automated driving system, a person's attention diverts away from the driving situation.*

4.4.2 Study design

The third user study was designed as a longer-term driving study to investigate the impact of prolonged system use on the development of trust. Participants used the automated driving system on a total of four study days. The study aimed at gaining additional insight about the development of trust after the first encounter with an automated driving system, and identifying the impact of a dedicated HMI over the long term. This section describes the procedure that was adopted to investigate the hypotheses mentioned above in a real-world driving study.

Participants

Participants of this user study were employees of Volkswagen Aktiengesellschaft and were recruited internally again. As compensation, participants received a gift voucher. 18 participants volunteered to take part in the user study (9 women and 9 men). They were 39.00 years old on average ($SD = 9.15$ years) and had an average mileage of 27 889 km of driving per year ($SD = 17\,101$ km), while they were not very experienced with driver assistance systems. When asked about their driving preferences, only two participants preferred being a passenger over driving themselves. The majority of participants liked the fun of driving as well as being in control of the situation. Only two drivers considered their driving skills to be worse than others', the rest of the drivers felt they had equal or better driving skills than average. None of the participants had experienced a self-driving vehicle before or was involved in the development of the functionality.

Test vehicle

As in the first user study, a prototype vehicle (Figure 4.39) with HAD functionality was used for the user study. The HAD system is described in Section 3.3.2 (see also Bendewald et al., 2015). Again, the vehicle was able to perform all aspects of the dynamic driving task on highway roads. Even if a human driver would not respond to a takeover request of the vehicle, the system would be able to perform a minimal risk maneuver by initiating a safe-stop as a fallback solution. In addition, a safety driver that was introduced as a technical support took place in the passenger seat and supervised the test drives to not put participants at risk at any time during the study. The automated driving system controlled the speed on its own and drove with a speed of 0 to 130 km/h (approx. 80 mph) while also controlling the distance to other vehicles and performing lane change maneuvers when appropriate.

Figure 4.39. Outside appearance of the concept vehicle used for the third user study.

The general interaction concept presented in Figure 3.6 was again applied in this study. The automated driving system could be activated or deactivated with two buttons on the steering wheel that had to be pressed simultaneously (5). Distinct sounds (6) and a unique color concept visible in the LED bar (3) were designed especially for automated driving and takeover requests. Different from the first study, however, information regarding the automated driving system was now displayed in the instrument cluster (1). Analysis of gaze data from the first user study had indicated that drivers expect system-related information to appear there.

Study design and independent factors

To investigate the impact of long-term as well as short-term effects of automated driving use on trust in the system, two repeated measures factors for a global and a local trend were included in the third user study. Also, the influence of system transparency during a longer system use was investigated and varied in the study with the help of the between-subjects factor HMI. The consideration of these factors resulted in a 2 (system transparency) x 4 (study day) x 3 (measurement) mixed factors study design.

Study days. The study's main focus was centered on the longer-term development of trust in an automated driving system over several days of practical experience with the system. This first factor, study days, was operationalized by four appointments for each participant within a period of five weeks (*global trend*). Study days were limited due to restricted access to the unique prototype vehicle. The objective of this longer-term familiarization was to minimize the impact of novelty of the situation on the results. To make sure every participant experienced the system during a longer time frame, appointments were made with a minimum time gap of two days. Only two participants needed to be rescheduled and drove on two subsequent days in a week.

Measurements. The second repeated measures factor was the time of measurement within each study day. One study day consisted of two drives with a short break in-between. Queries were carried out before the drive, during the break, and after the drive, to assess the *local trend* within the trials.

System transparency. The overall interaction concept of the prototype vehicle explained in Section 3.3.2 was utilized again. Two HMI versions that differed in their level of detail were designed for the study. The display content was derived from the concepts used in the first user study, but now differed more prominently regarding the provided information and was displayed in the instrument cluster. Furthermore, other parts of the interaction concept were also adjusted to provide a harmonious overall concept for each group. In the low-transparency condition (control group), the display content only informed the drivers that the system is active by showing a segmented circle in the dedicated turquoise color for automated driving (see Figure 4.40, left). Other information visualizing the system's actions were not provided. The LED bar implemented in the vehicle was turned off, not displaying any color during automated driving. Only during the takeover request, the LED bar turned orange and red. Furthermore, during the takeover request, a speech prompt consisting of a short note of a female voice saying “Piloted driving will be deactivated in 15 seconds.” and distinct sounds were audible. In the high-transparency version (experimental group), drivers as well received the information that the system is active with the help of the segmented turquoise circle. The circle was displayed in the lower part of the instrument cluster, while the main part of the instrument cluster was occupied by a visualization of the vehicle’s detected surroundings (see Figure 4.40, right). The location of the vehicle on the road, other road users, and intended lane changing maneuvers were visible for participants of the experimental group. In addition, during automated driving, the LED bar turned turquoise. For takeover, it turned orange and red and the female voice stated the longer request “Piloted driving will be deactivated in 15 seconds. Please take over steering.”, again underlined by distinct sounds.

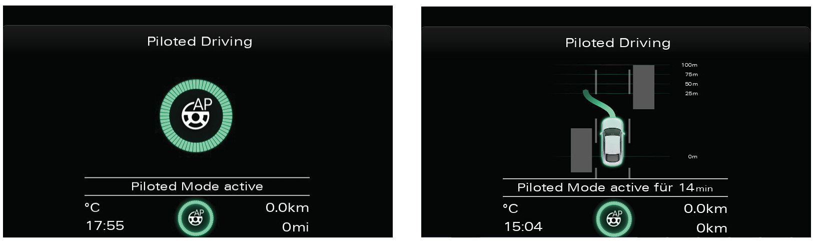


Figure 4.40. HMI versions used in the third user study. HMI version A (left) shows the status (active) of the HAD system, HMI version B (right) additionally shows the surroundings and the planned behavior of the vehicle.

As transparency of the automated driving system was a between-subjects factor, participants were randomly assigned to the control group or the experimental group. Each group saw one of the two HMI versions during their test drives (A – control group, B – experimental group). To maintain gender balance, gender was the only factor that was controlled during group assignment.

Dependent and mediating variables

To investigate how the presented independent variables influence trust in an automated driving system and other related variables, the following dependent subjective and objective variables were collected in the third user study. They were again collected using a windows tablet with a touch screen.

Subjective trust. As in the other user studies presented above, trust was assessed with a short version of a questionnaire by Muir (1989) that included four questions for predictability, dependability, faith, and overall trust in a system (Lee & Moray, 1992). Before and after each drive, participants had to rate their trust in the automated driving system on these items, again on a 15-point rating scale ranging from 1 = 'very low' to 15 = 'very high' (based on Heller, 1985).

Evaluation of the test drives. After each part of the test drives, single items regarding the perceived performance of the automated driving system, the nervousness during the automated drive, and the usefulness of the HMI indications needed to be rated on a 15-point scale ranging from 1 = 'very low' to 15 = 'very high' (based on Heller, 1985, see Appendix A.1.2, Table A.2).

Gaze behavior. Gaze behavior was measured again to objectify the driver's willingness to hand over control to the automated driving system. As in the two other user studies, the head-mounted Dikablis eye-tracking system (Version 3.0) was used to collect the data (Ergoneers GmbH, 2015). The areas of interest were the same as pictured in Figure 4.2. The attention focus of the driver was collected via the percentage of gazes on the street and on the instrument cluster. In this third user study, the change in gaze behavior over time was of special interest, as well as the drivers' interest in non-driving-related activities.

Distraction. Further, participants had the possibility to distract themselves from the driving task and watch a video in the center console display. Videos were available each run and could be watched at the participant's choice. The time until participants first engaged in this non-driving-related activity was measured with the help of the eye-tracking cameras. Also, the duration of use was gathered to find out how much participants dared to be distracted by a video. The duration was measured as the time a video was playing and as the sum of time periods (in s) participants focused their attention on the center console display.

Procedure

The study was set up as a real–driving study with the prototype concept vehicle described in Section 3.3.2. Figure 4.41 shows the setup within the vehicle.

The 18 participants were invited to take part in the study on four different days. Three appointments were carried out per day with three different participants. The sum of 72 appointments resulted in a complete study duration of five weeks. On their first study day, participants answered demographic questions and questions regarding their driving behavior. Participants were also introduced to the prototype vehicle and the study setting was explained (see Appendix A.1.1 for the instructions). On each day of participation, drivers experienced the HAD system (described

Figure 4.41. Setup of the third user study inside the automated driving test vehicle.

in Section 4.2) for approximately two hours (depending on traffic and weather conditions), resulting in a total of approximately eight hours of automated driving per person. Each study day, participants were asked to fill out a questionnaire in the beginning, during a short break after one hour of driving, and after the drive (see Figure 4.42).

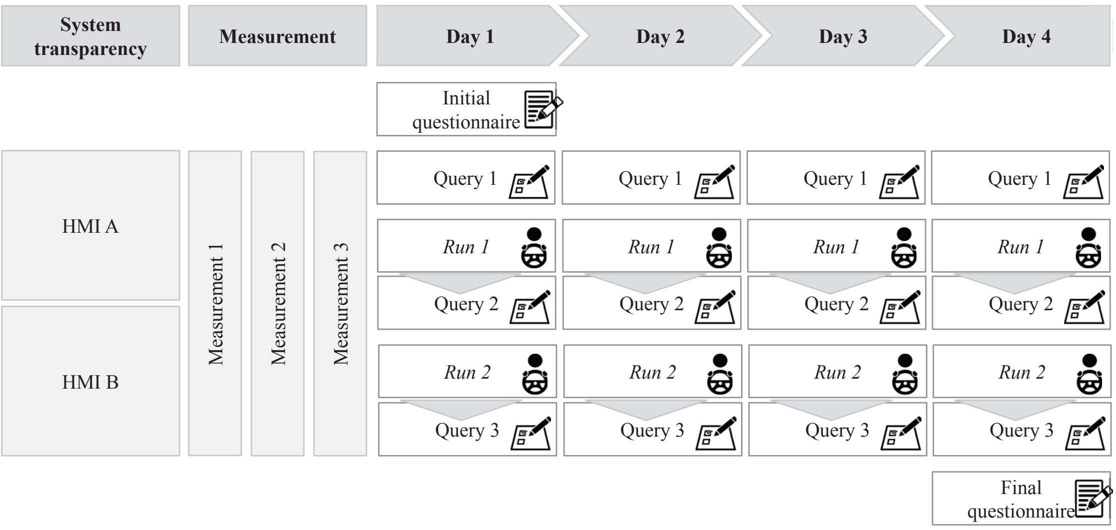


Figure 4.42. Procedure of the third user study.

All drives were attended by the experimenter on the back seat and the safety driver. Similar to the first real driving study, participants could activate the system with a simultaneous press of two dedicated buttons on the steering wheel, and were allowed to release the controls of the vehicle afterwards. Different from the arrangement of information in the cockpit for the first user study, system–related information was displayed in a more central position in the instrument cluster this time. The display in the center console was either displaying a navigation screen or videos, in case the participant started the video function with a press on the media button in the center console.

Of all 18 participants, 17 completed all four study days. One participant was unable to attend the last study day, resulting in 17 complete participations and one with only three study days.

4.4.3 Results

As in the other studies, IBM SPSS Statistics (version 19.0) was used for statistical analyses, and R (version 3.4.1) was used to create the accordant figures. The independent variables mentioned above were analyzed with inferential statistical methods according to the hypotheses described in Section 4.4.1. The dependent variables were assessed several times during the study, enabling analyses over time as well as overall analyses. They were checked for normal distribution with the Shapiro-Wilk test. With the help of an exploratory data analysis, extreme outliers were identified and removed from the analyses. Independent samples *t*-tests or repeated measures analyses of variances were conducted depending on the assumptions made beforehand. A result with an $\alpha = .05$ was defined as significant, and is presented in detail. Whenever necessary, the Greenhouse-Geisser correction was applied to correct for violation of sphericity. In Appendix A.2.3, Table A.16, an overview about the main variables, their descriptives, and their correlations can be found.

Hypothesis 1: Prolonged system use

In the first hypothesis, it was expected that depending on the duration of system use, the level of trust is adjusted. For initial system use, trust was expected to decline due to a positive automation bias before the first interaction. The high expectations regarding the automated driving system's performance was assumed to be relativized after the first use, creating the drop in trust. Afterwards, trust was expected to rise with prolonged system interaction.

To test the first sub-hypothesis, data of the first study day was analyzed by a 2 (HMI) x 3 (measurement) analysis of variance with repeated measures on the second factor. In this statistical model, a significant effect was found for the main factor measurement, $F(1.31, 20.99) = 8.63$, $p = .005$, $\eta_p^2 = .35$ (corrected with Greenhouse-Geisser). Paired comparisons revealed a significant difference between the first and the second measurement, $p_{1,2} < .001$. Unexpectedly, trust was lower before than after the first one-hour drive with the automated driving system. Figure 4.43 (left) shows the development of trust ratings within the first study day. The second sub-hypothesis referred to the remaining three study days. A 2 (HMI) x 3 (day) x 3 (measurement) model was analyzed. No overall main effect of the repeated measures factor day was found, $F(2,30) = 1.60$, $p = .219$, $\eta_p^2 = .10$. A tendency could be found for the factor measurement, $F(2,30) = 2.58$, $p = .092$, $\eta_p^2 = .15$. The mean trust ratings for study days 2 to 4 are depicted in Figure 4.43 (right), and show a small rise in trust especially from the second to the third measurement.

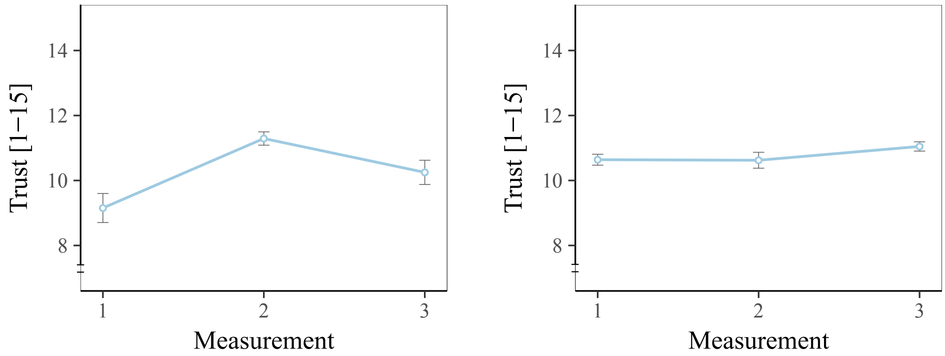


Figure 4.43. Mean trust ratings in the third user study for measurements on study day 1 (left) and for measurements on study days 2 to 4 (right). Error bars represent standard errors.

The chronological development of trust ratings in the third user study can be found in Figure 4.44. This complete presentation allows a visualization of both local and global effects over all study days. The trend of the local factor measurement can be seen in the figure. The figure also shows a high level of trust over all queries in general.

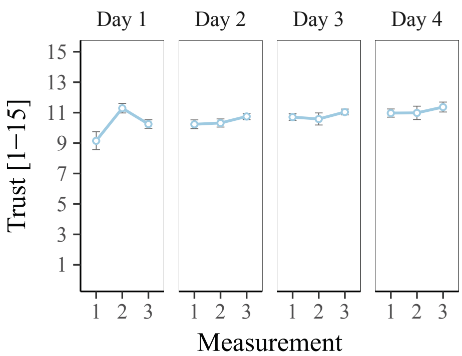


Figure 4.44. Mean trust ratings over the course of time in the third user study. Error bars represent standard errors.

The descriptive inspection of trust development over the course of the study shows a profoundly different development of trust at the first day of experiment compared with the following study days. To look at the different processes of trust in greater detail, Figure 4.45 shows the enlarged trust development at each study day on top of each other. At a closer look, the different quality of the first day's trust development becomes obvious. Different from the assumption, trust is comparatively low in the initial query (before the first drive), and rises to a higher level after the first encounter with the automated driving system. In the third rating, which took place at the end of the first study day and thus after the second drive, trust leveled off again. For the following study days, a different pattern becomes apparent. Always starting with the foregoing level of trust, ratings stay on the same level or are even diminished after the first part of the drive on the following days. Only in the third rating (after the second drive), trust rises slightly each day.

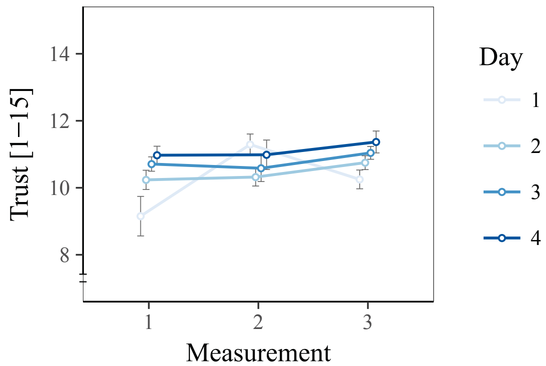


Figure 4.45. Enlarged mean trust ratings over the course of time in the third user study. Error bars represent standard errors.

Hypothesis 2: System transparency

Hypothesis 2 postulated a positive effect of system transparency on trust over time. The sub-hypotheses 2a and 2b distinguished between the initial interaction with the system and multiple practical experiences. In the beginning of the interaction with the system, no significant increase in trust was expected, as the foregoing studies did not find an effect of system transparency during short-term system use. However, high system transparency was assumed to display its effect in the long term, when system interaction gets more experienced. According to the sub-hypotheses, trust ratings for the two HMI versions were also analyzed separately for the first day and the subsequent study days.

For the first study day, the statistical model 2 (HMI) x 3 (measurement) did not show a significant main effect of the factor HMI, $F(1,16) = 1.44$, $p = .248$, $\eta_p^2 = .08$. Also, no interaction effect could be found for the factors measurement and HMI, $F(2,32) = 0.29$, $p = .747$, $\eta_p^2 = .02$. In Figure 4.46 (left) the trust development on the first study day depending on the HMI concept can be seen. Only the subsequent study days were included in the corresponding 2 (HMI) x 3 (day) x 3 (measurement) analysis. The repeated measures analysis of variance with the two within factors day and measurement and the between-subjects factor HMI showed no overall main effect of the between-subjects factor HMI, $F(1,15) = 1.60$, $p = .226$, $\eta_p^2 = .10$. The interaction between the factors measurement and HMI did not reach significance, $F(2,30) = 1.38$, $p = .268$, $\eta_p^2 = .08$. However, a tendency could be found for the interaction effect of the global factor day and HMI, $F(2,30) = 2.72$, $p = .082$, $\eta_p^2 = .15$. Paired comparisons showed a tendential difference both between study days 2 and 3, $p_{2,3} < .097$, and days 2 and 4, $p_{2,4} < .067$. This effect can be seen in Figure 4.46 (right).

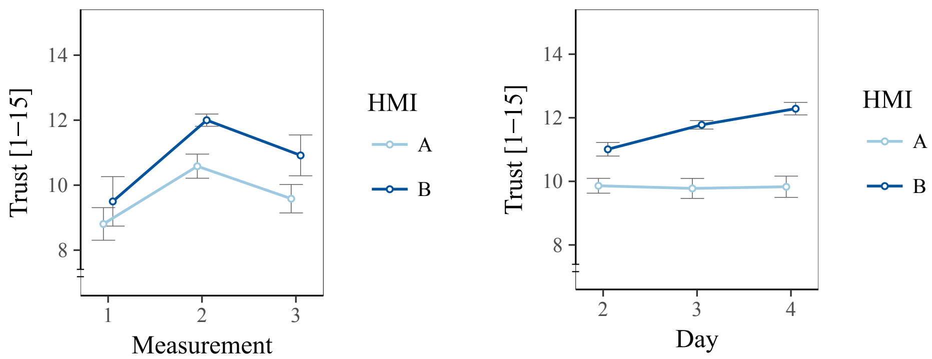


Figure 4.46. Mean trust ratings in the third user study depending on the HMI concept for measurements on study day 1 (left) and for measurements on study days 2 to 4 (right). Error bars represent standard errors.

Figure 4.47 shows the chronological development of trust ratings for the two HMI groups, to make local or global effects visible. On a descriptive level, it can be seen how trust ratings differed less in the beginning of the interaction (day 1) and split up more and more with each day (days 2 to 4), thus visualizing the tendential interaction effect between the global factor study days and the factor HMI. The more detailed HMI B received higher trust ratings during the later study days. Even more, one can see on a descriptive level that while with HMI A an adaptation phase with a decrease in trust was apparent each day, with HMI B this drop in trust was not as striking.

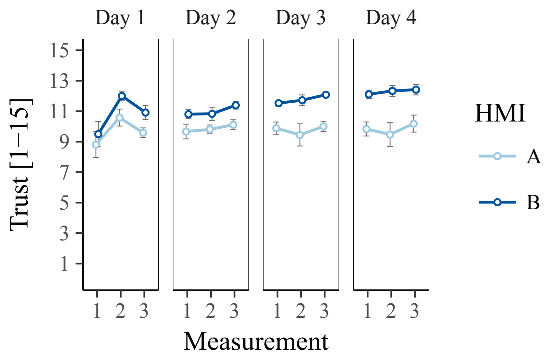


Figure 4.47. Mean trust ratings over the course of time depending on the HMI concept. Error bars represent standard errors.

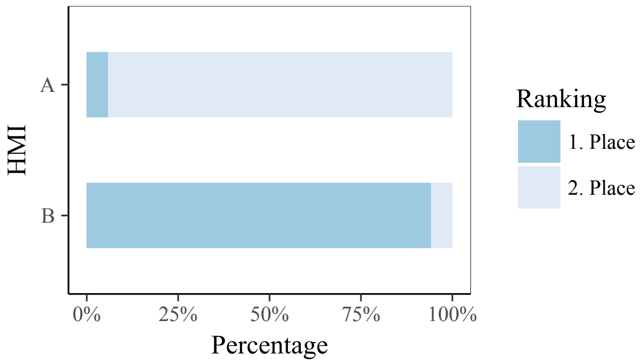


Figure 4.48. Results for subjective HMI preference in the third user study. The bar plot shows the percentage distribution of ratings for the two HMI concepts.

16 of the 17 participants who completed the study preferred the more detailed HMI B over HMI A (see Figure 4.48). Drivers significantly preferred getting detailed information about the automated drive even when they were not in charge of driving anymore, $\chi^2(1, N = 17) = 13.24$, $p < .001$.

Hypothesis 3: Trust and driver behavior

Lastly, the third hypothesis supposed that over time and with rising trust in the automated driving system, the driver's attention on the driving task will decrease. To check whether this assumption holds true, the attentional focus of the driver was observed, depending on the duration of system use (study days), on the transparency of the system (HMI), and on the level of trust the driver has developed. As an indication of the driver's diversion from the driving task, non-driving-related activities (videos) were allowed during automated driving phases.

Of all 18 participants, 14 made use of a non-driving-related activity during automated driving. The study situation might be a reason for so many participants exploiting the newly won freedom. Even hesitant participants were curious enough to try out a non-driving-related activity at least once during the four study days. It was observed when and for how long participants engaged in a non-driving-related activity during the test drives. To collect this data, two measures were taken into account. The duration a video was playing and the gazes in the direction of the non-driving-related activity (on the center console display) were measured.

The video duration was collected manually with the help of the eye tracking videos over the whole time. A repeated measures analysis of variance with the repeated factor day and the between-subjects factor HMI showed a significant difference in video duration depending on the HMI concept, $F(1,16) = 4.60$, $p = .048$, $\eta_p^2 = 0.22$, while no effect of the study day, $F(3,48) = 0.85$, $p = .476$, or an interaction could be found. Adding trust as a covariate (median split in high and low trust) the result was less striking, but the same tendency could still be found for the factor HMI, $F(1,15) = 4.13$, $p = .060$, $\eta_p^2 = 0.22$, while no effect could be found for trust, $F(1,15) = 0.13$, $p = .723$. Results hint at a difference depending on the HMI concept: drivers of the experimental group played videos for a longer period of time compared to the control group (see Figure 4.49). When analyzing the gaze percentage on the non-driving-related activity during HAD for the four study days, a similar result could be found. Here, a tendency in the same direction could be found with a repeated measures analysis of variance, $F(1,16) = 3.27$, $p = .089$, $\eta_p^2 = 0.17$, showing that the percentage of gazes on the center console screen depended on the HMI concept (see Figure 4.50).

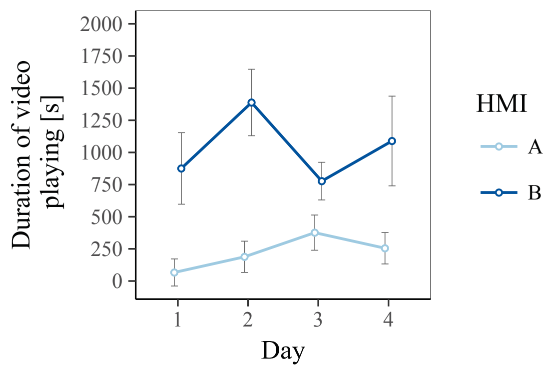


Figure 4.49. Mean duration of video playing during HAD over the course of the third user study depending on the HMI concept. Error bars represent standard errors.

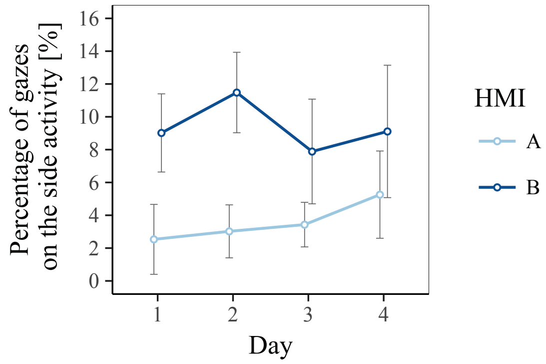


Figure 4.50. Mean percentage of gazes on the non-driving-related activity during HAD depending on the HMI concept. Error bars represent standard errors.

Again, no effect of study days, $F(2.05, 32.72) = 0.22$, $p = .807$, or an interaction effect was found. Adding trust as a covariate also resulted in a less conspicuous result (HMI: $F(1, 15) = 2.88$, $p = .110$; trust: $F(1, 15) = 0.29$, $p = .599$).

Participants seemed to focus their attention more on the center console display with the video when driving with the more detailed HMI concept. Figure 4.51 visualizes this result by showing an overview over the distribution of gazes on the areas of interest.

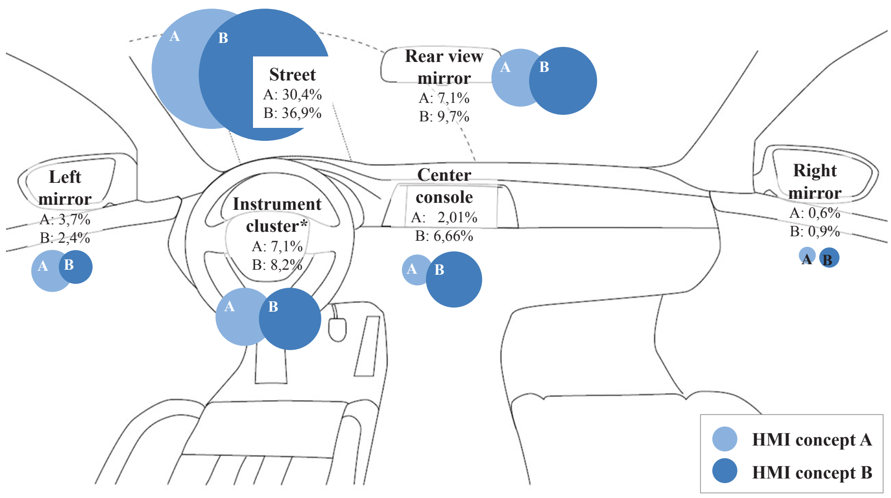


Figure 4.51. Percentage of gazes (attention ratio) depending on the HMI concept (A or B). The size of the circles is relative to the percentage of gazes in this area.

Taking the results of video and gaze duration analysis together, drivers engaged in non-driving-related activities longer and diverted their attention more to the center console screen when getting detailed information about the automated drive. This result was independent of their level of trust in the system. Also, no difference in non-driving-related activity use could be found over time.

An additional analysis was conducted to look at the time to first use of a non-driving-related activity. With an average time of 147 minutes (approximately 2.5 hours) since activating the automated driving system for the first time, people who saw the simple HMI A took longer until engaging in a non-driving-related activity compared to people who drove with the more detailed HMI B, who took only 69 minutes (a little more than 1 hour) on average until diverting from the driving scene. However, an independent *t*-tests for the two HMI groups as well as more and less trusting people (median split in high and low trust) did not reveal a significant effect of the HMI, $Welch's\ t(1.07) = 0.57$, $p = .667$, or trust, $t(8) = 0.43$, $p = .677$. *Welch's t*-test was used due to unequal variances of the two HMI groups, as Levene's test revealed ($p = .013$).

4.4.4 Implications

The objective of the last user study was to identify changes in the level of trust in the automated driving system over a longer period of time, while using a certain HMI concept to inform the driver about the automated drive.

Findings. Trust rose when using the automated driving system for the first time. This result does not fit with sub-hypothesis 1a, which had postulated a different result based on the previous user studies. Trust rose slightly within the study days, while no global trend over the course of the four study days could be found. The tendential local trend is consistent with sub-hypothesis 1b. From the results, one can assume that trust stays stable when no interaction with the automated system takes place. It seems that an adaptation process starts whenever another interaction with the automated driving system takes place. This process leads to a step-wise development of trust in the system during each encounter. On a descriptive level, it was found that the first system contact follows a different pattern than the rest of the encounters with the system. Going along with this result, a variation of the HMI did not have an effect on trust ratings in the beginning of the interaction. Sub-hypothesis 2a can thus be accepted. In sub-hypothesis 2b, it was postulated that the HMI will have an effect trust during prolonged system interaction. A tendential significance of the HMI x day interaction effect partially confirmed this hypothesis. Trust ratings for the two HMI groups differed more with each day of system interaction, and the more detailed HMI B seems to have led to a rise in trust over time. In hypothesis 3, drivers were expected to divert their attentional focus away from the driving situation. Results regarding the factors time, system transparency and trust found confirmation for sub-hypothesis 3b. Drivers did not change their gaze focus or their use of non-driving-related activities over time or with higher trust, but they engaged in non-driving-related activities longer and diverted their attention more to the center console screen when getting detailed information about the automated drive.

Different phases of trust development were identified with the study. Results of the initial study day were somewhat surprising. In the previous user studies a decrease in trust after the first drive was found, confirming the presence of an automation bias before the first drive. This time, trust increased after the first drive, indicating that people had more trust after experiencing the automated driving system for the first time. This rise in trust could be a positive reminiscence of a special experience. Still, trust decreased after the second drive. Possibly, only after the second drive the expectations were brought down. The decrease of trust in the beginning of each subsequent study day (after the initial drive) could be explained by a variety of effects. An explanation for trust to be decreasing could be the route of the study. As the study always started at the same place, the order of route sections could not be varied. An order effect could be the consequence, if for example the first part of the route was perceived as less trustworthy than the second part. The effect of decreasing trust could also be explained by a familiarization phase. Drivers may need some time to gain trust in the system again after not using it for a while. This adaptation seems to get more difficult with each time, indicating a repeated, stable adaptation phase. Between the test drives, it was shown by the study results that trust is a relatively stable construct that is not degraded over a short time, when no interaction with the system takes place. Within the scope of this study, a time frame of a few days was taken into account. Potentially, trust will decrease after a longer time (weeks or months) without system interaction. Interestingly, when comparing the curves for the different HMI concepts, the patterns of trust grow more different with each day. The tendency in the interaction effect of HMI x time supports the descriptive observations, indicating that the factor HMI could have a noteworthy influence on trust over time. Due to a relatively high effect size, the effect may still be seen as relevant. In the third and fourth study days, with the detailed HMI concept trust seemed to increase. Based on this observation, it seems that with a more transparent system, the adaptation process mentioned before is less relevant and can be overcome to create a constant growth of trust in the system with ongoing interaction. Since the interaction effect was only tendentially significant, further investigations must follow.

Figure 4.52 shows an interpretation of the descriptive findings. It is suggested from the descriptive data that trust development in an automated driving system can be divided into different phases. The phase of *initial contact* seems to be fundamentally different from the later phases, showing a rise in trust in the very first interaction, which is then diminished when interaction with the automated driving system continues. In the following phase of trust development, drivers' trust seems to follow an *adaptation process* (phase I). While trust stays stable when no interaction with the system takes place, it takes some time for drivers to get familiar with the

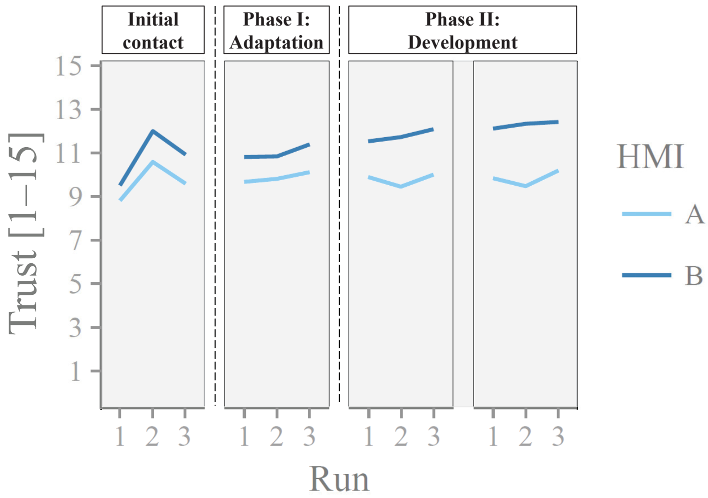


Figure 4.52. Suggested phases of trust development for automated driving, derived from the results of the user studies.

system when interacting with it again. A sharp bend in trust can thus be found in the following interactions with the automated vehicle, that is resolved after a while when trust rises again to the level it had before the new interaction. Results hint at a *development phase* (phase II) for interaction with a transparent system. It could be interpreted that with an adequate HMI concept, the adaptation phase can be overcome and trust is not diminished during the following interactions, leading to a continuous rise in trust with each interaction. These interpretations based on descriptive data need to be verified in further studies. Finally, it can be speculated that trust development will change later on, when permanently using the automated driving system. This last phase of trust development might lead to a ceiling effect with trust not rising any further. However, this real driving study covered a few weeks of occasional usage, not a permanent use over month and years. It thus can only be speculated how trust in the automated driving system will develop further.

The evaluation of the HMI concepts supplied evidence that the need for information did not decline over time and with prolonged system use. The more detailed HMI version was preferred for longer use by nearly all participants of the study. For the trust development phases addressed in this study, the information given by the HMI should thus not be reduced. Other information strategies could be considered for potential later phases, when trust is established (e.g., varying the amount of indications given in a certain situation).

Limitations. In this user study on trust in automated driving across several experiences, some aspects could not be controlled due to its setting as a real driving study. For example, weather and traffic situations were different for each participant and each study day. Heavy rain hindered the activation of the automated driving system in some occasions, traffic congestions had to be circumvented, and construction sites were built on the road. Thus, participants experienced the automated driving system in slightly different environments and on slightly different stretches of the road. However, a minimum of 30 minutes of HAD per drive (1 hour per day) was always assured.

As already emphasized, the duration of system use may have been too short. Four days of driving do not represent a long time compared to the expertise drivers develop driving manually in their day-to-day experiences. Yet, compared to solely one initial contact with the system in the previous studies, several sessions of automated driving still offered the possibility of getting to know the automated driving system better. It can be assumed that trust in the automated driving system will reach its peak at some point, but this point could not be found within the course of this study.

The complex long-term study design made it difficult to achieve a high number of cases for the study, as participation was time-consuming and a regular attendance of participants was crucial. Conducting the study with 18 participants resulted in a study duration of approximately five weeks. Still, for a more profound statistical analysis, a larger sample would be needed. The small sample resulted in less powerful statistical tests and a lower probability to detect smaller effects. However, this also strengthens the found significant effects and tendencies. As in the previous studies, participants stemmed from a possibly biased group of employees of the Volkswagen Aktiengesellschaft. For this reason, any generalization of insights must be made with caution.

Caused by the same starting point each day, the order of the route was always the same, and no randomization of the road sections was possible. This way, sequence effects could not be excluded completely. However, the road sections were very similar, and an effect on the results was thus unlikely.

As in the first user study, the study situation with an observer and a researcher in the car and a pair of eye tracking glasses on the head can influence the way people feel during the test drive. Also, some participants may have noticed the possibilities of the safety driver to control the vehicle. This had a positive influence on trust ratings in the first user study. The phases of trust development might thus be accelerated due to the study setting and could take longer when people are alone in the vehicle. A further indication of this assumption is the reported level of trust in the automated driving system, which was generally very high in the study.

During the course of the five-week study, news of a deadly accident involving a semi-automated vehicle (Tesla Model S) spread and attracted widespread attention. The vehicle was driving in semi-automated mode (Level 2 automation, see Table 2.2), when it did not detect a crossing tractor-trailer and crashed into it. The information about this accident was relevant for the participants of the study, and was discussed by them during the test drives. As this unforeseen event was uncontrollable, participants received this information at different points of familiarity with the test vehicle. It was examined on a descriptive level if trust ratings changed after the event, but no dramatic differences could be found. A reason might be that every participant had experienced the test system at least once before the event, so that they were able to differentiate between the systems and did not apply the accident to the test vehicle at hand.

Conclusion. This third user study was designed to address research limitations found in literature and in the foregoing user studies, by assessing trust development in highly automated vehicles in a real environment and over a longer time frame. Furthermore, it completes the research results gained in this work and puts them into perspective.

In summary, it can be concluded from the study that a transparent HMI concept in a central cockpit position may help drivers to maintain and even increase their trust in the automated driving system over time. Also, it was found that drivers engage in non-driving-related activities longer when having the detailed information in front of them. Different phases of trust development were distinguished with the help of the study. Thereby, the study could show how different factors shape trust over time, which can be helpful to design transparent automated driving systems in the future. Finally, the study was also a proof of feasibility, showing that although cumbersome, longer-term studies can be necessary to discover human processes like trust development.

5 Discussion

This chapter discusses the results obtained in the three user studies presented in Chapter 4. In Section 5.1, an overview is given over the main results of these user studies. Section 5.2 provides their evaluation. It puts the results into an overall context and derives implications of the newly acquired knowledge in Section 5.2.1, and discusses limits of the research at hand in Section 5.2.2. Finally, Section 5.3 discusses the need for further research.

5.1 Overview of results

The main objective of this work is to examine how trust in a HAD system develops and how it can be supported. To identify important influencing factors and test a specific HMI concept for automated vehicles, three user studies were conducted:

- *Study 1* – Real-driving study: Investigation of the influence of personality characteristics and attitudes on trust in an automated driving system.
- *Study 2* – Simulator study: Variation of system reliability to explore if trust can be maintained with the help of system experience or system transparency.
- *Study 3* – Longer-term real-driving study: Development of trust in automated driving across multiple practical experiences.

With the help of a prototype vehicle in real traffic (studies 1 and 3) and a simulated setting (study 2), initial trust in the system was observed as well as trust development over a longer time frame. Relevant factors for trust were derived from existing trust models. Accordingly, the influence of personality factors, system characteristics like transparency (HMI concept), and system experience were included in the investigation.

Study 1: Individual differences in trusting an automated vehicle. The first user study provided research results on individual differences influencing initial trust in automated driving with new insights due to its setting in a real driving environment. The analysis of parameters assessed in the study suggested an influence on trust by demographic factors like age as well as by personality characteristics like technology acceptance and desire for control. Furthermore, driver's experience of the performance of the automated driving system was found to be crucial for trust development. Experiencing one situation with a system boundary or low system reliability did not have a strong impact on trust, but several situations of this kind diminished trust profoundly. The results of this study strengthen the link between trust and human behavior with an automated system, showing trust as a behaviorally relevant construct for automated driving. High trust in the system led to a shift of attention allocation and more engagement in non-driving-related activities. Results thus imply that both system characteristics and personality factors influence the level of trust in automated vehicles. The implications drawn from the results of this user study in real traffic environment were used to guide the subsequent user studies. The HMI concepts were updated and further hypotheses were formulated.

Study 2: The importance of system reliability. In the simulator user study, the consequences of system boundaries were observed in detail. Refining the results of the first user study regarding the high importance of perceived system performance, it can now be specified that the experience of system limits will only lead to a low valuation of system performance if it can actually bring about a dangerous situation. Other system limits that may cause unusual driving behavior or unnecessary maneuvers were not considered a bad driving performance by the participants of the study. It was found that system experience can help to maintain trust when system performance is perceived as low. This result complies with the results of Sanchez et al. (2011), who showed that the impact of low system reliability depends on the level of experience with the system. While system transparency did not have an effect on trust, a higher transparency of the automated driving system was considered useful and was thus preferred. To verify whether system transparency will become more important over time, another study in real driving context was planned.

Study 3: Beyond initial trust—Trust across multiple practical experiences. The longer-term real-driving study was designed to investigate trust beyond the initial reactions addressed in former research on trust in automated driving. In this study, trust was observed for a longer period of system use, while differentiating between two different levels of system transparency. While an effect of system transparency on system trust could not be found in the first two user studies, the results of the prolonged observation hinted at a difference over time. After the initial phase of system interaction, trust in the system only rose when transparency of the system was high. With no further system information, the descriptive observation showed that driver's trust underwent an adaptation process at each system encounter, resulting in a volatile level of trust. It can thus be suspected that the factor system transparency gains importance during system use. Other factors, such as certain personality characteristics and perceived system performance, are more important in the beginning of system interaction, as the results of the first and second user study show.

The next section will evaluate these results and will embed them into other research done on that topic.

5.2 Evaluation of studies

The focus of this work was on expanding knowledge about how trust in an automated driving system develops and how it can be encouraged. Within this work, three main research questions were formulated and tried to answer with the help of the user studies:

- What impacting factors and correlates for trust exist regarding the interaction with an automated driving system?
- Can system transparency engender a pertinent level of trust in an automated driving system, even in the event of a system limit?
- How does trust in an automated driving system evolve and what are the connotations of trust in different stages of system use?

5.2.1 Study findings and implications

To answer the research questions and examine how trust in HAD can be influenced, user studies were conducted in simulated and real driving settings and initial interaction with an automated driving system as well as prolonged system use were observed.

The research presented in this work makes real-world data on trust in automation available for the novel context of automated driving (first and third user study). This raises the generalizability of the presented results and has the potential to verify former research results on this topic. By definition, trust is a concept relevant in situations of uncertainty and vulnerability (see Section 2.2.1, definition by Lee & See, 2004), and measurements can thus be assumed to be of higher validity under real driving conditions. Nonetheless, some scientific issues make it necessary to investigate research questions in the more secure and standardized setting of a simulator. As the second user study investigated the topic of system boundaries, it was conducted in a driving simulator.

A result all three user studies had in common is the high level of trust in the automated driving system in general, as has been found in other research before (e.g., Eimler & Geisler, 2015). This implies that to investigate trust in this context in detail, a very sensitive means of measurement is required.

Effects of system boundaries Performance of the automated driving system was identified as the most relevant factor influencing trust in the system (going along with result of e.g., Hancock et al., 2011). While this is not surprising, interesting further details were found in the second user study. Not only was it shown that experience with the automated driving system can help in the event of system boundaries (e.g., impossible or unnecessary driving maneuvers), but also that there are differences in trust ratings depending on the type of the system boundary. So-called 'false alarms', in the context of automated driving for example unnecessary lane changes due to a misinterpretation of the driving situation, do not diminish driver's trust in the system. 'Misses', where the driving system does not detect an object, reduce trust considerably. This is unusual compared to other research, where false alarms of a decision aid system appeared "to be more damaging to overall performance than misses" (Dixon et al., 2007, p. 564). The difference can be explained with the consequences for the driving experience. Unnecessary maneuvers do not put the driver in a dangerous situation, while a missed object might well end in an unintended situation, where maneuvers need to be aborted abruptly. No system errors are expected in HAD due to redundancy of sensors, but nonetheless implications for the design of the system's driving behavior are apparent. Developers of automated driving systems need to consider these results when designing a function. Different from other research recommendations, it might be advisable to adjust the threshold of sensor detection to be less conservative for automated driving systems compared to advisory systems. With that, the driving system would drive more carefully in case it detected something (true of not). Misses could be reduced that way, which were found to be worse for trust development in the automated driving system than false alarms.

Effects of system transparency. To pursue the research questions, an approach with a dedicated HMI concept for automated driving was chosen. Former research had found evidence for the importance of system transparency with the use of automated systems (e.g., Billings, 1996; Christoffersen & Woods, 2002; Sarter & Woods, 1995). Interestingly, the first two user studies could not confirm these findings for automated driving systems. In these settings, trust was not influenced by the level of detail the information about the automated system had. One potential reason is the small difference between the HMI concepts and their depiction of information. With a longer-term observation of trust development and the insights of the last user study, it became clear that a reason for the lacking effect in the first studies could be the short-term observation of interaction with the system. This initial contact was found to be profoundly different from further trust development. For initial contact, the personality factors acceptance of technology and desire for control were important next to the actual performance of the system. The transparency of the system (the HMI concept) was not relevant for initial trust. Over time, this proportion may change in favor of the HMI concept, which rose in influence. Differences in the stages of trust development had been found by Muir and Moray (1996) before, where faith was found to be a good predictor of initial use of automation rather than of later stages of trust development. The findings also go along with Madsen and Gregor (2000) and Lee and See (2004). They describe that initial trust is more related to affective processes of the human mind, while subsequent trust is strongly influenced by analytic processing of the machine's abilities (Merritt & Ilgen, 2008). Potentially, this implies that for the first encounter with an automated driving system, a different approach is necessary—regarding the HMI concept as well as regarding the performance and driving style of the automated driving system. A special tutoring might be advisable for drivers using an automated driving system for the first time.

Effects of system transparency over time. The results can explain why other research, for example, a similar approach by Kleen et al. (2014), did not reveal significant effects of the carefully designed HMI concept on system trust. Such results mostly refer to the first contact with the system. This does not imply that drivers do not need to be informed about the driving behavior of the automated vehicle, as one could assume. It does merely show that people need the information later on, when they have become acquainted with the whole situation and can concentrate on other details apart from the vehicle's driving behavior itself. Most drivers are not receptive during the very first interaction with such a novel system. The fact that drivers need between 1 hour and 2.5 hours on average until they turn to a non-driving-related activity also fits in well with this finding. Based on these results, one can assume that the whole process of trust development needs to be supported by an HMI concept on an ongoing basis.

This development of trust in an automated driving vehicle over time is depicted in Figure 4.44. Due to the study setting, the time frames given in this figure might underestimate the time it really takes to develop trust in a vehicle that drives itself. Due to drivers not being alone in the vehicle, being observed as part of a study, and sitting in a much tested vehicle, the times found in the user study might show a time lapse of what would happen under normal circumstances. In a situation where drivers are driving alone in an automated vehicle and are the only ones able to intervene if necessary, the process of getting confident with the vehicle can be assumed to take longer.

In this research's time frame, no ceiling effect could be found. Trust had not reached the maximum of the scale and was still growing after a few days with the support of the detailed HMI concept. Possibly, there will be a turning point later on, when users have become acquainted with an automated driving system over several month or even years. At that point it might not be necessary to provide detailed information about the system anymore, because it is already understood and trusted. However, this point was not reached in the studies due to limitations in time and availability of the test vehicle, and it remains an open question when this moment will be reached. As the trust development process depends, among others, on personality characteristics, it will most likely be a point in time that is individual for each driver.

Model of trust in automated driving. This research furthermore enhances the knowledge about the theoretical relationship between trust and influencing factors relevant to the context of automated driving. Important parameters influencing trust in a HAD system have been identified in this work and were linked to a certain time period within the process of trust development. According to the results of the user studies, there is a substantial difference between initial and learned trust. Initial trust behavior is mainly influenced by attitudes and personality characteristics, especially acceptance of technology and desire for control. Trust development is then influenced most of all by the perceived performance of the automated driving system, but potentially also by the transparency of the system over prolonged system use. The working model of trust in HAD that was based on Hoff and Bashir (2015) as well as Lee and See (2004) (presented in Section 3.1) was revised on the basis of these research results, to also reflect the different phases of trust development. Figure 5.1 shows how dispositional trust, which is based mostly on human characteristics, impacts on initial trust behavior (*initial contact phase*). After experiencing the automated driving system, system and environmental characteristics become most relevant for learned and situational trust, with performance of the system being important right from the beginning of the interaction (*phase I*) and the HMI presumably gaining importance over time (*phase II*). Thus, while in the beginning of the interaction the user can only base his reliance decision on affect-based trust, analytic processes gain relevance with growing system experience (as Madsen and Gregor (2000) and Lee and See (2004) suggested). The connection of influencing factors with the temporal aspect is of both methodological and practical importance. The temporal differentiation can be helpful for the interpretation of research results as well as when trying to influence and change trust. In both cases, it is necessary to understand which phase of trust is examined in order to derive an adequate approach. When addressing initial trust, one might be able to influence trust by changing the general attitude towards automated systems or the performance of the system (if possible), but not necessarily by providing a detailed HMI concept. When looking at later trust phases, the HMI potentially will become more influential and can be used as a means of altering trust. The consequences regarding HMI design are discussed further on in the next paragraph.

Recommendations for an automated driving HMI concept. A dedicated HMI concept for automated driving is needed to give users the chance to get to know the new technology. This finding is not trivial, as has been shown above, and neither is the design of an appropriate HMI

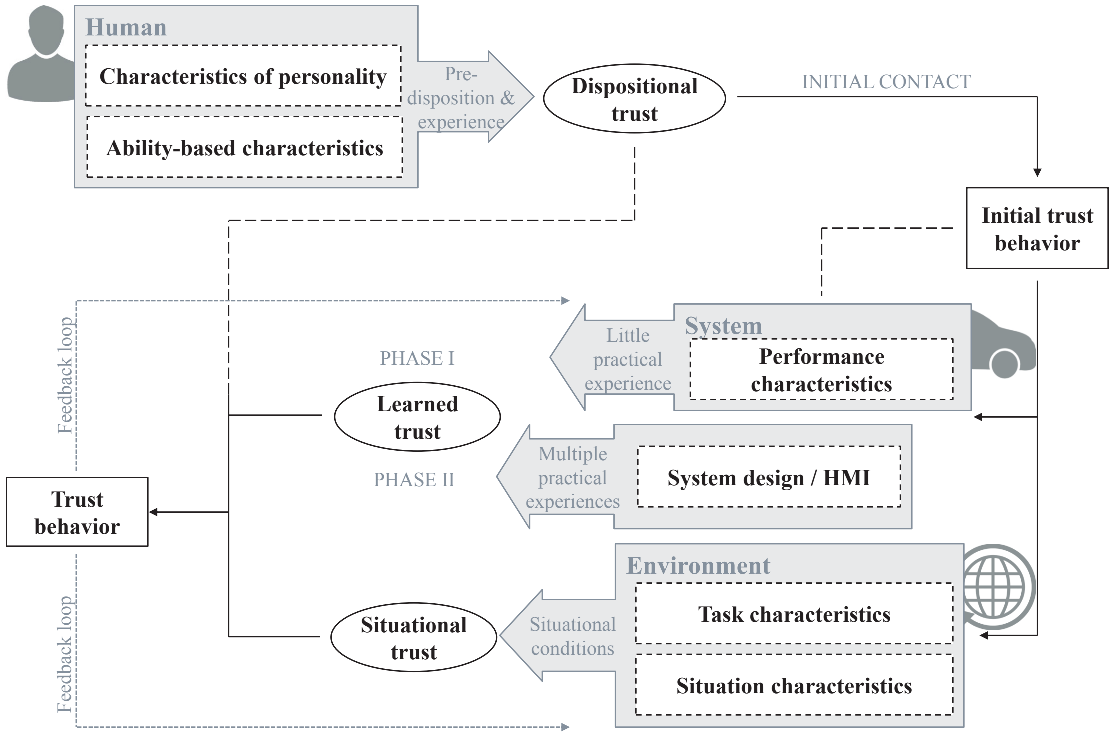


Figure 5.1. Enhanced model of trust in HAD, based on the results of the user studies (originally based on the models of Hoff and Bashir (2015) as well as Lee and See (2004)).

concept. It needs to be chosen wisely which information should be displayed to actually increase the driver's confidence in the system. Displaying the raw data of the vehicle's sensors might not always promote trust in the system. On the contrary, if false alarms or misses of the detection are made visible, it might even unsettle the driver. Yet, for example for semi-automated driving systems, it might be necessary to indeed display exactly this information to enable the driver to decide whether the system is capable of handling the situation or not. This way, appropriate trust could be achieved in lower automation levels with the help of an HMI concept. For HAD, system boundaries are expected to be detected and announced by the system. An HMI concept with detailed information about what the sensors detect and what actions the system is planning to undertake (describing, explaining, and predicting information) can help to refine the driver's mental model of the system. This was also suggested by Rouse et al. (1992) (depicted in Figure 2.10). The preservation of the driver's mental model is technically not necessary while driving highly automated—yet the information supports trust in the same time, as research results have shown. Thus, though the reason for informing the driver has changed, it is still recommended to provide information about the system's behavior.

The illustration in Figure 5.2 summarizes the phases of trust development for automated driving that have become evident in the three user studies, describes the essential aspects of each trust level, and indicates according HMI requirements. The phases of initial contact, adaptation, and development were apparent in the studies and recommendations regarding the design of the user interaction and information can be drawn. Phase III with a very high system trust and less need for information could not be assumed from the user studies, and can thus only be speculated on. So far, it can hence be recommended to address the phases of trust development before established trust is (potentially) reached and support the user of automated driving systems accordingly. Research in the automotive area also supports the findings gained in this work. A survey of the Massachusetts Institute of Technology investigated the consumer preference to learn about the technology implemented in an automated vehicle (Abraham et al., 2016). 39% of all participants would like the car to teach them how to use the technology, next to the vehicle's manual (59%), or websites (38%).

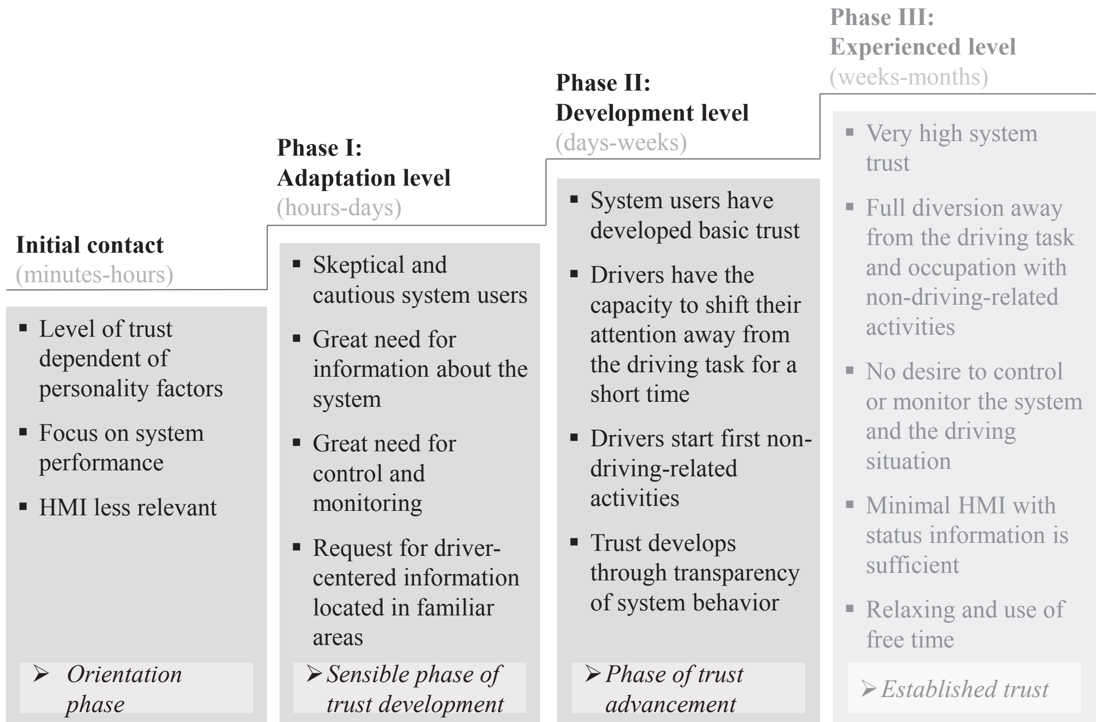


Figure 5.2. Stages of trust development for automated driving.

Design guidelines applicable to trust in automated driving were collected in Section 2.3.2 (see Table 2.6). They can help to take user trust into account while designing automated driving systems. Nonetheless, they need to be translated into concrete implementations, the presented

HMI concept being one possible solution. It has been pointed out that the required level of system transparency within the process of trust development can depend on several factors, e.g., the driver's personality and the level of experience with the system. It can thus be advisable to design the HMI configurable for the (expert) user, to enable him to decide when detailed information can be reduced or taken away. Future, more intelligent systems might even be able to learn what information they should provide and to adjust system transparency on their own. An example of how a configurable concept could look like was created by Peugeot with their concept car 'Instinct' (Dillet, 2017). It provides two modes of automated driving for the user: 'autonomous sharp' and 'autonomous soft'. The driver can choose a mode and thereby not only decide upon the driving style, but also upon the degree of detail in the HMI concept. While the sharp version favors the quickest route and informs the driver about all driving-related processes, the soft mode shall provide a smooth ride and does only show the selected mode.

First legal frameworks are developed to restrict and coordinate the use of automated vehicles on public roads. They also make the HMI of an automated vehicle a subject of discussion. In their Federal Automated Vehicles Policy, the NHTSA (2016, p. 22-23) asks for indications that the highly automated vehicle is at a minimum

1. functioning properly,
2. currently engaged in automated driving mode,
3. currently 'unavailable' for automated driving,
4. experiencing a malfunction with the HAD system, and
5. requesting control transition from the HAD system to the operator.

Also, the legislative proposal for the German Road Transport Law includes the demand that the vehicle has to communicate the need to take over to the driver visually, acoustically, or haptically (Deutscher Bundestag, 2017). However, the research at hand shows that the need for information is much greater on the side of the users. First providers of automated driving (e.g., Uber in Pittsburgh and San Francisco) are paying attention to customer needs as well. In their vehicles (that still have a human driver inside as a fallback), Uber installed a screen that explains concerns like vehicle speed and the purpose of all sensors on the roof. Also, it tells the current driving mode, speed, route, and displays the vehicle with its surroundings as understood by the sensors (Davies, 2016).

It can be concluded that, despite not being in control anymore, drivers still want and need to get information about the automated driving system. The system's perception and cognition, its actions, decisions, and processes during a highly automated drive can be important—not because they need to monitor the system, but because they need to understand the system to trust it. If we cease our efforts of designing transparent driving systems under the misconception that it will get obsolete with self-driving technology, this might have serious implications. If we do not support drivers with the new technology, we risk people not daring to use the technology, which would mean not exploiting the expected benefits of automating transportation, including the gain in traffic safety.

5.2.2 Limitations

The user studies were planned and conducted with great care. Nonetheless, they have some limitations that can compromise their validity in some aspects.

Sample of participants. Participants of the user studies conducted in this research were part of an internal pool of employees or former employees of the Volkswagen Aktiengesellschaft. It is possible that this selective sample of participants is not completely representative, as employees of a car manufacturer might be more interested in automotive topics than others and potentially have a higher interest in and acceptance of technology in general. To keep this potential bias as small as possible, only people not involved in the development of assisted or automated driving functions were allowed to participate. Participants had no specific driving expertise (e.g., through a prototype training), to be as representative as possible. Still, it is possible that the results of the studies presented in this work are biased because of a technologically adept sample. Therefore, any interpretation of results needs to be made with caution.

Real driving and simulator settings. As was already mentioned in Section 5.2.1, both real driving and simulator settings have their advantages and drawbacks. Especially for trust in the context of automated driving, it is getting more and more important to test in real settings. Of course, when using an actual highly automated vehicle and driving in real traffic, external disturbing factors are rising in influence. Weather, road, and traffic conditions cannot be held constant. Furthermore, even in this realistic setting, the unusual test situation cannot be circumvented. In the presented real driving studies, a safety driver and an experimenter were always sitting in the prototype vehicle with the participant. The presence of other passengers can influence the experience of participants, potentially making them feel more secure during the test drive than they normally would. Others might feel more distracted, or might not want to engage in non-driving-related activities when feeling observed. A fully natural behavior will only occur outside of a test situation. The advantages and disadvantages of simulator settings have been summarized by De Winter et al. (2012). The more artificial situation in a simulated environment reduces the validity of the results compared to a real-world study. For testing automated driving, the missing lateral acceleration as a performance indicator is a major drawback (at least for a static simulator). Nonetheless, some research questions require the safe and controlled environment of a simulator (e.g., to test a system in dangerous situations). The results of the second user study emerged from a simulated environment and their comparability with the results of the real driving studies are thus limited.

Study duration. In the third user study, it was attempted to investigate long-term trust in an automated driving system. The limited availability of the prototype vehicle resulted in a maximum of four days of driving experience for each participant. It became clear that the approximately six to eight hours of driving in HAD mode per participant did not suffice to map the entire process of trust development. Trust was still rising with the transparent HMI concept, indicating that the process of trust development was not completed yet. It can be expected that trust will settle at a relatively stable level at some point, but this point in time was not reached within the scope of this research.

Statistical analysis. The number of cases for each user study was limited by external factors, especially by time constraints due to the limited availability of the prototype vehicle and the driving simulator. The small number of participants, particularly of the real driving studies, resulted in a low power of the statistical tests. This can explain some tendencies found in the data that did not reach significance. However, it also strengthens the significant differences that were found in the data. The questionnaire used to measure trust in the system (designed by Muir (1989) and shortened by Lee and Moray (1992)) was helpful for assessing trust in an easy to understand and not too time-consuming way. In case a more detailed analysis with different subscales of trust is needed, other questionnaires might be more useful.

5.3 Further research

This work provides insights into the topic of trust development in automated vehicles. The results can be useful for future design of automated driving systems and their interaction concepts. Yet, the research also raises further questions that need to be part of future efforts.

The working model for trust in automated driving described in Section 3.1 should not be considered to be exhaustive. It is a conglomeration of current research and transfers the most important factors influencing trust into the context of automated driving. Additional influencing factors are conceivable and should be addressed in subsequent research.

This work focuses on the level of high automation in driving systems. The HAD system can be assumed to function reliably and implies that the user can completely trust the system to do so. Thus, overtrust is not an issue during HAD—takeover situations being the sole exception. They are the only situations in which the driver must be attentive again. During the highly automated drive, trust can be as high as possible without disadvantages. For this reason, *calibration* of trust was not made a subject of discussion here. However, it is a highly relevant issue for semi-automated driving and needs to be analyzed in detail for that context.

During HAD, the driver is allowed to engage in a non-driving-related activity while the automated driving system is active. A lot of research is already conducted to determine which non-driving-related activities can be allowed and which should be legally restricted due to a too long takeover process (e.g., sleeping). Driver monitoring and the design of the vehicle's takeover request can play an important role in this context. Messages could be adjusted to the driver's current state to create an appropriate warning (i.e., louder sounds if the driver is reading, less prominent notifications if the driver is already looking at the road). Analyses of driver's responsiveness depending on the non-driving-related activity and different takeover requests can provide further insights. A new approach could even involve gamification—"the use of game design elements in non-game contexts" (Deterding, Khaled, Nacke, & Dixon, 2011, p. 1)—to encourage drivers to stay alert and attentive (Burkert, 2016). It might also be a way to give drivers back the fun of driving some are afraid to lose.

Longer-term investigations regarding trust and other implications of system use need to be conducted. The research presented here provides a first approach and can serve as a feasibility study. It already became apparent that the time frame used in the third user study might still have been too short to see the whole picture. Possibly, there will be a ceiling effect with trust not getting any higher. This might result in a reduced need for information at some point, where the HMI concept could show reduced information. When trust development will reach this point is yet to be determined.

It is concluded from the research at hand that transparent systems are crucial for trust in and use of novel automated driving technology and should be pursued further in research. The HMI concept designed in this work constitutes one possible solution to increase transparency of an automated driving system. There have been other attempts to give information regarding such a system, for example displaying uncertainty of the automated driving system (Beller et al., 2013; Helldin et al., 2013). Uncertainty information helped drivers in a simulator study of Helldin et al. (2013) to prepare for a takeover situation after a failure of the semi-automated driving system. Also, this group of drivers reported lower, more calibrated trust in the system compared to the control group without uncertainty information. Other information or other modalities could potentially be used to enhance transparency of an automated driving system. Furthermore, the pre-tests of this work revealed the relevance of the HMI location. Performance of the vehicle was rated differently when information was shown in a HUD. The HMI location was not pursued in the main studies, but should be considered further in future studies.

An HMI concept could furthermore be supplemented by other approaches to familiarize users with automated driving systems, for example an integration in drivers' education. The training of humans interacting with automation has been suggested by Atoyan et al. (2006) before (see Table 2.6 in Section 2.3.2). The differentiation between different levels of automated driving as well as the interaction with an accordant system could be trained and accompanied by a technical support in the beginning. This approach could enhance trust in the new technology further and should be part of future research.

The HAD vehicle's driving style was identified in a pre-study as a potential factor influencing trust. In the course of this work, the factor was not addressed any further, mainly because of the technical effort of implementing different driving styles into the prototype vehicle. Research on this factor shows that it is relevant for acceptance of automated driving (Hartwich, Beggiato, & Krems, 2018).

When designing interaction concepts for automated driving, one should not only address the inside of the vehicle (driver and passengers), but also the surrounding interaction partners like pedestrians, bicyclists, and other vehicles. When eye contact cannot be used, a very important means of communication in traffic is missing and other ways to communicate need to be found. Especially in mixed traffic, where automated vehicles drive next to human drivers, this will become a major issue. Concepts need to be found to make these encounters as safe as possible. On the one hand, road traffic regulations may establish designated lanes for automated vehicles. On the other hand, developers of automated vehicles may equip the cars with means to communicate to the outside world as well (e.g., visual indicators, sounds). Concepts for this outward communication need to be created and tested. Concepts that improve the outward appearance and communication might even be able to positively influence public trust in the novel technology.

6 Conclusions

This thesis provides research results on trust in HAD. For the first time, these are based on real-world data and prolonged system use. The research results demonstrate the importance of an adequate HMI in the context of trust in automated driving. Furthermore, they can foster an active debate about how to take user trust into account in the design of automated driving systems. The main outcomes of this thesis are valuable both for further scientific research as well as for practical application.

Conclusion: Level of trust. In general, a high level of trust in the automated driving system was found. This indicates that drivers will likely be willing to test the new functionality. While this faith in technology is good in that it supports the development towards automated driving, over-trust still poses a risk during lower driving automation levels. However, results show that drivers do not blindly trust the automated driving system right from the start—they need to be convinced that the technology is ready to serve as a chauffeur. The first challenge is thus to get the driver out of the loop rather than back in. Even when initial trust is high, it is a fragile concept that can be destroyed easily, and is difficult to rebuild once lost.

Conclusion: Influencing factors. Perceived system performance is the one variable that predicted trust in the system best. Early in the process of trust development, personality characteristics predicted trust as well. Later in the process, system transparency had an effect on the development of system trust. As the results of the user studies showed, drivers prefer a more transparent system and gain more trust in it over time when detailed system information is given. Thus, even though drivers are not actively engaged in the driving task anymore, it is still recommended to provide them with detailed information about the automated driving system to enhance system understanding and long-term trust.

Conclusion: Theoretical insights. The thesis examined a model of trust in HAD that is based on the models of Lee and See (2004) and Hoff and Bashir (2015). The data in this work supported the trust model in the specific context of automated driving. The model differentiates between initial trust in a system (after a short, superficial contact with a system) and learned trust in the system (after becoming familiar with the system during prolonged system use). This temporal differentiation helps to identify what kind of trust is measured. Moreover, it determines influencing factors relevant for a specific type of trust. With the help of the trust model, this thesis identified the stages of trust that can be influenced by the HMI.

Conclusion: Recommendations for HMI design. The recommendations given in this work help to accommodate the user's trust during the development of automated driving systems. This work gives advice on when transparent information is most relevant for drivers and proposes a detailed HMI concept. Some projects and enterprises already implement this strategy of informing the driver even without him being in charge of driving (e.g., Tesla, Uber, Peugeot). The results of this thesis encourage such an approach, especially during trust formation. When system trust is high enough or has reached its maximum, the level of detail of the HMI concept may be reduced—however, this moment will be highly individual. Developers may thus want to consider a configurable HMI concept that enables drivers to reduce information when they do not need it anymore.

Directions for future research. Automated driving may sooner or later become a part of our daily lives. Until then, questions that still remain unaddressed after this work need to be answered. A longer time frame of system use needs to be discussed with regards to system trust. Also, a general higher familiarity with such systems due to a wider dissemination of them can be taken into account in future research. The presented results can to some extent also be transferred to fully automated driving. Still, it might be a different feeling sitting in a vehicle that is expected to cope with some traffic situations in comparison to a vehicle that is supposed to be capable of handling every situation. By proving that real-driving studies and observations across multiple experiences are possible in this field of research, this thesis broadens current knowledge and encourages further research.

In sum, it can be noted that with the careful design of an in-car HMI concept for automated driving functionality, we can meet the challenges of the novel technology and strengthen the usage intention. Creating a transparent system is crucial for trust in and use of the novel technology of HAD.

List of Figures

List of Tables

Nomenclature

ABS	Anti-lock braking system
ACC	Adaptive cruise control
AdaptIVe	Research project “Automated driving applications and technologies for intelligent vehicles”
AOI	Area of interest
APT	Argument-based probabilistic trust model
BASt	Bundesanstalt für Straßenwesen, in English Federal Highway Research Institute
BMWi	Bundesministerium für Wirtschaft und Energie, in English Federal Ministry for Economic Affairs and Energy
CMS	Collision mitigation system
DARPA	Defense Advanced Research Projects Agency
EPoSS	European Technology Platform on Smart Systems Integration
EU	European Union
HAD	Highly automated driving
HAVEit	Research Project “Highly automated vehicles for intelligent transport”
HMI	Human-machine interface
HUD	Head-up display
interACT	Research project “Designing cooperative interaction of automated vehicles with other road users in mixed traffic environments”
InteractIVe	Research project “Accident avoidance by active intervention for intelligent vehicles”

IPIP	International Personality Item Pool
Ko-HAF	Research project “Cooperative Highly Automated Driving”
LED	Light-emitting diode
LKA	Lane-keeping assistant
LOA	level of automation
NASA	National Aeronautics and Space Administration
NEO PI-R	Revised NEO Personality Inventory
NHTSA	National Highway Traffic Safety Administration
OEM	Original Equipment Manufacturer
SA	Situation awareness
SAE	Society of Automobile Engineers
VDA	Verband der Automobilindustrie, in English: German Association of the Automotive Industry
VTD	Virtual Test Drive (simulator software)

Bibliography

Abraham, H., Lee, C., Brady, S., Fitzgerald, C., Mehler, B., Reimer, B., & Coughlin, J. F. (2016). *Autonomous Vehicles, Trust, and Driving Alternatives: A survey of consumer preferences* (Tech. Rep.). Cambridge, MA: Massachusetts Institute of Technology.

Adams, B. D., Bruyn, L. E., Houde, S., & Angelopoulus, P. (2003). *Trust in automated systems - Literature review* (Tech. Rep.). Toronto, Canada: Defence Research and Development Canada.

Alberdi, E., Strigini, L., Povyakalo, A. A., & Ayton, P. (2009). Why Are People's Decisions Sometimes Worse with Computer Support? *Computer Safety, Reliability, and Security*, *5775*, 18–31. doi: 10.1007/978-3-642-04468-7

Albert, M., Lange, A., Schmidt, A., Wimmer, M., & Bengler, K. (2015). Automated Driving: Assessment of Interaction Concepts Under Real Driving Conditions. *Procedia Manufacturing*, *3*, 2832–2839. doi: 10.1016/j.promfg.2015.07.767

America's Independent Electric Light and Power Companies. (1956). Power companies build for your new electric living. *LIFE Magazine*, *40*(5), 3.

Atoyan, H., Duquet, J.-R., & Robert, J.-M. (2006). Trust in New Decision Aid Systems. *Proceedings of the International Conference on Association Francophone d'Interaction Homme-Machine*, 115–122. doi: 10.1145/1132736.1132751

Audi AG. (2012). *Audi Blog: Piloted Driving.* Retrieved 08.03.2018, from `http://blog.audi.de/tag/piloted-driving/`

Audi AG. (2016). *Audi A7 piloted driving concept.* Retrieved 08.03.2018, from `https://www.audi-mediacenter.com/de/fotos/album/audi-a7-piloted-driving-concept-646`

Baddeley, A. D. (1996). *Working Memory*. Oxford, U.K.: Clarendon.

Bahner, J. E. (2008). *Übersteigertes Vertrauen in Automation: Der Einfluss von Fehlererfahrungen auf Complacency und Automation Bias* (Doctoral dissertation). University of Technology Berlin, Germany.

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779. doi: 10.1016/0005-1098(83)90046-8

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York: W. H. Freeman.

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

Beggiato, M., Hartwich, F., Schleinitz, K., Krems, J., Othersen, I., & Petermann-Stock, I. (2015). What would drivers like to know during automated driving? Information needs at different levels of automation. *7. Tagung Fahrerassistenz*. doi: 10.13140/RG.2.1.2462.6007

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. doi: 10.1016/j.trf.2012.12.006

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.

Bendewald, L., Glaser, E., Petermann-Stock, I., & Stephan, A. (2015). „Jack" – A holistic approach of designing a human machine interface for highly-automated driving. In *VDI Berichte – International Congress Electronics in Vehicles* (Vol. 17, pp. 453–467). Düsseldorf: VDI-Verlag GmbH.

Billings, C. E. (1991). *Human-Centered Aircraft Automation: A Concept and Guidelines* (Tech. Rep.). Moffett Field, CA: National Aeronautics and Space Administration.

Billings, C. E. (1996). *Human-Centered Aviation Automation: Principles and Guidelines* (Tech. Rep.). Moffett Field, CA: National Aeronautics and Space Administration.

Billings, C. E., & Cheaney, E. S. (1981). *Information Transfer Problems in the Aviation System* (Tech. Rep.). National Aeronautics and Space Administration.

Breznitz, S. (1984). *Cry Wolf: The Psychology of False Alarms*. Englewood Hills, N. J.: Lawrence Erlbaum Associates.

Broadbent, D. E. (1958). *Perception and communication*. Elmsford, NY: Pergamon Press, Inc.

Brunswik, E. (1952). *The conceptual framework of psychology*. Oxford, England: University of Chicago Press.

Buld, S., Krüger, H.-P., Hoffman, S., Kaussner, A., Tietze, H., & Totzke, I. (2002). *Wirkungen von Assistenz und Automation auf Fahrerzustand und Fahrsicherheit* (Tech. Rep.). Würzburg, Germany: Interdisziplinäres Zentrum für Verkehrswissenschaften.

Buld, S., Tietze, H., & Krüger, H.-P. (2005). Auswirkungen von Teilautomation auf das Fahren. In M. Maurer & C. Stiller (Eds.), *Fahrerassistenzsysteme mit maschineller Wahrnehmung* (pp. 161–187). Berlin, Heidelberg: Springer-Verlag.

Burger, J. M., & Cooper, H. M. (1979). The desirability of control. *Motivation and Emotion*, *3*(4), 381–393.

Burkert, A. (2016). *Gamescom Congress 2016: Zum Spielen ins Auto*. Retrieved 30.04.2017, from `https://www.springerprofessional.de/automatisiertes-fahren/ergonomie---hmi/gamescom-congress-2016-zum-spielen-ins-auto/10606218`

Cacciabue, P. C. (2004). Elements of Human-Machine Systems. In P. C. Cacciabue (Ed.), *Guide to Applying Human Factors Methods* (pp. 9–47). London: Springer-Verlag.

Cai, H., & Lin, Y. (2010). Tuning Trust Using Cognitive Cues for Better Human-Machine Collaboration. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(28), 2437–2441. doi: 10.1177/154193121005402816

California Legislative Counsel. (2012). *Senate Bill 1289, Chapter 570, Division 16.6 "Autonomous Vehicles"*.

Chancey, E. T., Bliss, J. P., Liechty, M., & Proaps, A. B. (2015). False Alarms vs. Misses: Subjective Trust as a Mediator between Reliability and Alarm Reaction Measures. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *59*(1), 647–651. doi: 10.1177/1541931215591141

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, pp. 1–12). New York: Elsevier.

Cohen, M. S., Parasuraman, R., & Freeman, J. T. (1998). Trust in Decision Aids: A Model and its Training Implications. *Proceedings of the Command and Control Research and Technology Symposium*, 1–37.

Davies, A. (2016). *We Take a Ride in the Self-Driving Uber Now Roaming Pittsburgh.* Retrieved 22.04.2017, from `https://www.wired.com/2016/09/self-driving-autonomous-uber-pittsburgh/#slide-1`

De Craen, S., Twisk, D. A. M., Hagenzieker, M. P., Elffers, H., & Brookhuis, K. A. (2011). Do young novice drivers overestimate their driving skills more than experienced drivers? Different methods lead to different conclusions. *Accident Analysis and Prevention*, *43*, 1660–1665.

De Winter, J. C. F., Van Leeuwen, P. M., & Happee, R. (2012). Advantages and Disadvantages of Driving Simulators: A Discussion. In A. J. Spink, F. Grieco, O. E. Krips, Loijens, L. W. S., Noldus, L. P. J. J., & P. H. Zimmerman (Eds.), *Proceedings of Measuring Behavior* (pp. 47–50). Utrecht, Netherlands.

Desai, M. (2012). *Modeling trust to improve human-robot interaction* (Doctoral dissertation). University, Lowell, Massachusetts.

Desai, M., Medvedev, M., Vázquez, M., McSheehy, S., Gadea-Omelchenko, S., Bruggeman, C., … Yanco, H. A. (2012). Effects of Changing Reliability on Trust of Robot Systems. *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 73–80. doi: 10.1145/2157689.2157702

Deterding, S., Khaled, R., Nacke, L. E., & Dixon, D. (2011). Gamification: Toward a Definition. In *CHI 2011 Gamification Workshop Proceedings* (pp. 6–9).

Deutscher Bundestag. (2017). Gesetzesentwurf der Bundesregierung: Entwurf eines ... Gesetzes zur Änderung des Straßenverkehrsgesetzes: Drucksache 18/11300.

De Vries, P. (2004). *Trust in Systems: Effects of Direct and Indirect Information* (Doctoral dissertation). University of Technology Eindhoven, Netherlands.

De Waard, D., Van der Hulst, M., Hoedemaeker, M., & Brookhuis, K. A. (1999). Driver behavior in an emergency situation in the automated highway system. *Transportation Human Factors*, *1*(1), 67–82.

Dijksterhuis, C., Stuiver, A., Mulder, B., Brookhuis, K. A., & de Waard, D. (2012). An Adaptive Driver Support System: User Experiences and Driving Performance in a Simulator. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(5), 772–785. doi: 10.1177/0018720811430502

Dillet, R. (2017). *Peugeot's Instinct concept car has active driving and autonomous modes.* Retrieved 22.04.2017, from `https://techcrunch.com/2017/02/27/peugeots-instinct-concept-car-has-active-driving-and-autonomous-modes/`

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the Independence of Compliance and Reliance: Are Automation False Alarms Worse Than Misses? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(4), 564–572. doi: 10.1518/001872007X215656

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. doi: 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. doi: 10.1518/0018720024494856

Eimler, S. C., & Geisler, S. (2015). Zur Akzeptanz Autonomen Fahrens – Eine A-Priori Studie. In A. Weisbecker, M. Burmester, & A. Schmidt (Eds.), *Mensch und Computer 2015 Workshopband* (pp. 533–540). Stuttgart: Oldenbourg Wissenschaftsverlag.

Endsley, M. R. (1988). Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *32*, 97–101.

Endsley, M. R. (1993). Situation Awareness and Workload: Flip Sides of the Same Coin. *Proceedings of the International Symposium on Aviation Psychology*, 906–911.

Endsley, M. R. (1995). Measurement of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 65–84. doi: 10.1518/001872095779049499

Endsley, M. R., Bolte, B., & Jones, D. G. (2003). *Designing for Situation Awareness: An Approach to User-Centered Design*. London: Taylor & Francis.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, *42*(3), 462–492.

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. doi: 10.1518/001872095779064555

Ergoneers GmbH. (2015). *D-Lab.* Retrieved 22.04.2017, from `http://www.ergoneers.com/`

European Technology Platform on Smart Systems Integration. (2015). *European Roadmap: Smart Systems for Automated Driving* (Tech. Rep.). Berlin, Germany: European Technology Platform on Smart Systems Integration.

Fitts, P. M. (1951). *Human Engineering for an Effective Air Navigation and Traffic-Control System* (Tech. Rep.). Washington, DC: Ohio State University, Research Foundation.

Flight Safety Foundation. (2005). Increased Reliance on Automation May Weaken Pilots' Skills for Managing System Failures. *Human Factors & Aviation Medicine*, *52*(2), 1–4.

Gasser, T. M. (2012). Legal issues of Driver Assistance Systems and Autonomous Driving. In A. Eskandarian (Ed.), *Handbook of Intelligent Vehicles* (pp. 1519–1535). London: Springer-Verlag.

Gebhardt, W. A., & Brosschot, J. F. (2002). Desirability of Control: Psychometric Properties and Relationships with Locus of Control, Personality, Coping, and Mental and Somatic Complaints in Three Dutch Samples. *European Journal of Personality*, *16*(6), 423–438.

German Federal Ministry of Transport and Digital Infrastructure. (2015). *Strategie automatisiertes und vernetztes Fahren* (Tech. Rep.). Germany: Federal Ministry of Transport and Digital Infrastructure.

German Federal Ministry of Transport and Digital Infrastructure. (2017). *Ethics Commission: Automated and Connected Driving* (Tech. Rep.). Germany: Federal Ministry of Transport and Digital Infrastructure.

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

Gold, C., Damböck, D., Bengler, K., & Lorenz, L. (2013). Partially Automated Driving as a Fallback Level of High Automation. *6. Tagung Fahrerassistenz. Der Weg zum automatischen Fahren*.

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. doi: 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.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., De Visser, E. J., & Parasuraman, R. (2011). A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(5), 517–527. doi: 10.1177/0018720811417254

Harris, M. (2015). *Uber could be first to test completely driverless cars in public.* Retrieved 09.03.2017, from `http://spectrum.ieee.org/cars-that-think/transportation/self-driving/uber-could-be-first-to-test-completely-driverless-cars-in-public`

Hartwich, F., Beggiato, M., & Krems, J. F. (2018). Driving comfort, enjoyment and acceptance of automated driving: Effects of drivers' age and driving style familiarity. *Ergonomics*, *61*(8), 1017–1032.

Heimgärtner, R. (2007). Cultural Adaptive Human Machine Interaction in Driver Information and -Assistance Systems. In T. Paul-Stueve (Ed.), *Mensch & Computer Workshopband* (pp. 37–40). Weimar: Verlag der Bauhaus-Universität.

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

Heller, O. (1985). Hörfeldaudiometrie mit dem Verfahren der Kategorienunterteilung (KU). *Psychologische Beiträge*, *27*, 478–493.

Herczeg, M. (2014). Design for Responsibility: Die Rolle des Operateurs zwischen Automation und Verantwortung. In M. Grandt & S. Schmerwitz (Eds.), *Der Mensch zwischen Automatisierung, Kompetenz und Verantwortung* (pp. 1–16). Bonn: Deutsche Gesellschaft für Luft- und Raumfahrt (DGLR).

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. *International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, *8*, 337–343.

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.

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. doi: 10.1177/0018720814547570

Hutchins, E. (1995). *Cognition in the Wild.* Cambridge, MA: The MIT Press.

interACT. (2017). *Designing cooperative interaction of automated vehicles with other road users in mixed traffic environments.* Retrieved 07.07.2018, from `https://www.interact-roadautomation.eu/`

Itoh, M. (2012). Toward overtrust-free advanced driver assistance systems. *Cognition, Technology & Work*, *14*(1), 51–60.

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.

Johannsen, G. (1993). *Mensch-Maschine-Systeme*. Berlin, Heidelberg: Springer-Verlag. doi: 10.1007/978-3-642-46785-1

Johnson, J. A. (2006). *Short form for the IPIP-NEO (International Personality Item Pool representation of the NEO PI-R).* Retrieved from `http://www.personal.psu.edu/faculty/j/5/j5j/IPIP/ipipneo120.htm`

Kaber, D. B., & Endsley, M. R. (1997). Out-of-the-Loop Performance Problems and the Use of Intermediate Levels of Automation for Improved Control System Functioning and Safety. *Process Safety Progress*, *16*(3), 126–131.

Karrer, K., Glaser, C., Clemens, C., & Bruder, C. (2009). Technikaffinität erfassen – der Fragebogen TA-EG. In A. Lichtenstein, C. Stößel, & C. Clemens (Eds.), *Der Mensch im Mittelpunkt technischer Systeme: 8. Berliner Werkstatt Mensch-Maschine-Systeme* (pp. 196–201). Düsseldorf: VDI-Verlag GmbH.

Kaufmann, C., Risser, R., Geven, A., & Sefelin, R. (2008). Effects of simultaneous multi-modal warnings and traffic information on driver behavior. In C. Brusque (Ed.), *Proceedings of European Conference on Human Centred Design for Intelligent Transport Systems* (pp. 33–42). Lyon: Humanist publications.

Kazi, T. A., Stanton, N. A., Walker, G. H., & Young, M. S. (2007). Designer Driving: Drivers' Conceptual Models and level of trust in Adaptive-Cruise-Control. *International Journal of Vehicle Design*, *45*(3), 339–360.

Kleen, A., Ricknäs, D., Schäfer, H., Manelli, M., Bendewald, L., & Selmarker, A. (2014). Automatische Fahrfunktionen und Mobile Devices: HMI-Konzept "Active Frame": Darstellung von system- und umweltbezogenen Informationen auf Tablet PCs während der automatischen Fahrt. *Proceedings of the 30. VDI Fahrerassistenz und Integrierte Sicherheit*, 199–214.

Knapp, R. K., & Vardaman, J. J. (1991). Response To an Automated Function Failure Cue: an Operational Measure of Complacency. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *35*(2), 112–115. doi: 10.1518/107118191786755931

Koustanai, A., Cavallo, V., Delhomme, P., & Mas, A. (2012). Simulator Training With a

Forward Collision Warning System: Effects on Driver-System Interactions and Driver Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(5), 709–721. Retrieved from `http://hfs.sagepub.com/cgi/doi/10.1177/0018720812441796` doi: 10.1177/0018720812441796

Kyriakidis, M., Happee, R., & De Winter, J. C. F. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour*, *F 32*, 127–140. doi: 10.1016/j.trf.2015.04.014

Large, D. R., Burnett, G. E., Morris, A., Muthumani, A., & Matthias, R. (2018). A Longitudinal Simulator Study to Explore Drivers' Behaviour During Highly-Automated Driving. In N. A. Stanton (Ed.), *Advances in Human Aspects of Transportation.* Springer International Publishing AG.

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

Lee, J. D., & Moray, N. (1994). Trust, self-confidence and operators' adaptation to automation. *International Journal of Man-Machine Studies*, *40*(1), 153–184.

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.

Lewandowsky, S., Mundy, M., & Tan, G. (2000). The Dynamics of Trust: Comparing Humans to Automation. *Journal of Experimental Psychology: Applied*, *6*(2), 104–123.

Llinas, J., Bisantz, A. M., Drury, C. G., Seong, Y., & Jian, J.-Y. (1998). *Studies and Analyses of Aided Adversarial Decision Making: Phase 2: Research on Human Trust in Automation* (Tech. Rep.). Air Force Research Laboratory, State University of New York.

Luhmann, N. (2000). Familiarity, Confidence, Trust: Problems and Alternatives. In D. Gambetta (Ed.), *Trust: Making and Breaking Cooperative Relations* (pp. 94–107). Oxford, U.K.: Basil Blackwell Ltd.

Ma, R. (2005). *The Effect of In-Vehicle Automation and Reliability on Driver Situation Awareness and Trust* (Doctoral dissertation, State University, North Carolina). doi: 10.1016/j.ergon.2005.04.002

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. doi: 10.1080/14639220500337708

Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2003). Automation Failures on Tasks Easily Performed by Operators Undermines Trust in Automation Aids. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *47*(3), 335–339.

Madsen, M., & Gregor, S. (2000). Measuring Human-Computer Trust. *Proceedings of the Australasian Conference on Information Systems*, 6–8.

Manzey, D. H., 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. doi: 10.1177/1555343411433844

Masalonis, A. J., Duley, J. A., Galster, S. M., Castano, D. J., Metzger, U., & Parasuraman, R. (1998). Air traffic controller trust in a conflict probe during free flight. *Proceedings of the 42nd Annual Meeting of Human Factors and Ergonomics Society*, *42*(23), 1601.

Masalonis, A. J., & Parasuraman, R. (1999). Trust as a Construct for Evaluation of Automated Aids: Past and Future Theory and Research. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(3), 184–187. doi: 10.1177/154193129904300312

Maurer, M., Gerdes, J. C., Lenz, B., & Winner, H. (2015). *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte*. Berlin, Heidelberg: Springer-Verlag.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, *20*(3), 709–734.

McBride, M., & Morgan, S. (2010). *Trust Calibration for Automated Decision Aids* (Tech. Rep.). North Carolina: Institute for Homeland Security Solutions.

McGuirl, J. M., & Sarter, N. B. (2003). Supporting Trust Calibration in Automated Decision Aids Through the Presentation of System-generated, Dynamic Confidence Information. *Proceedings of the International Symposium on Aviation Psychology*, *2*, 791–796.

Merat, N. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*(27), 274–282. doi: 10.1016/j.trf.2014.09.005

Merat, N., & Jamson, A. H. (2009). How do drivers behave in a highly automated car? *Proceedings of the International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 514–521.

Merritt, S. M. (2011). Affective Processes in Human-Automation Interactions. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(4), 356–370. doi: 10.1177/0018720811411912

Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, J. D. (2013). I Trust It, but I Don't Know Why: Effects of Implicit Attitudes Toward Automation on Trust in an Automated System. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*(3), 520–534. doi: 10.1177/0018720812465081

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 Hu-*

man Factors and Ergonomics Society, *50*(2), 194–210. doi: 10.1518/001872008X288574

Metzger, U., & Parasuraman, R. (2001). Automation-Related "Complacency": Theory, Empirical Data, and Design Implications. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *45*(4), 463–467.

Miller, C. A. (2005). Trust in Adaptive Automation: The Role of Etiquette in Tuning Trust via Analogic and Affective Methods. In *Proceedings of the International Conference on Augmented Cognition* (pp. 1–9).

Miller, C. A., & Parasuraman, R. (2007). Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(1), 57–75.

Moray, N. (2000). Are Observers Ever Really Complacent When Monitoring Automated Systems? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(6), 592–595. doi: 10.1177/154193120004400613

Moray, N., & Inagaki, T. (1999). Laboratory studies of trust between humans and machines in automated systems. *Transactions of the Institute of Measurement and Control*, *21*(4-5), 203–211. doi: 10.1177/014233129902100408

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied*, *6*(1), 44–58.

Mosier, K. L., & Skitka, L. J. (1996). Human Decision Makers and Automated Decision Aids: Made for Each Other? In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 201–220). Hillsdale, N.J.: Lawrence Erlbaum.

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation Bias: Decision Making and Performance in High-Tech Cockpits. *The International Journal of Aviation Psychology*, *8*(1), 47–63.

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. doi: 10.1016/S0020-7373(87)80013-5

Muir, B. M. (1989). *Operators' trust in and use of automatic controllers in a supervisory process control task* (Doctoral dissertation). University of Toronto, Canada.

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. doi: 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. doi:

10.1080/00140139608964474

National Highway Traffic Safety Administration. (2015). *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey: Traffic Safety Facts, Crash Stats. Report No. DOT HS 812 115* (Tech. Rep.). Washington, DC.

National Highway Traffic Safety Administration. (2016). *Federal Automated Vehicles Policy: Accelerating the Next Revolution In Roadway Safety* (Tech. Rep.). Washington, DC.

Niederée, U., & Vollrath, M. (2009). Fahrerassistenzsysteme der Zukunft: Fährt da der Mensch noch mit? In VDI Wissensforum GmbH (Ed.), *5. VDI-Tagung: Der Fahrer im 21. Jahrhundert.* Düsseldorf: VDI-Verlag GmbH.

Norman, D. A. (1989). *The Design of Everyday Things*. New York, NY: Doubleday/Currency.

Norman, D. A. (1990). The problem of automation: inappropriate feedback and interaction, not over-automation. *Philosophical transactions of the Royal Society of London*(B), 585–593. doi: 10.1093/acprof:oso/9780198521914.003.0014

Numan, J. H. (1998). *Knowledge-Based Systems as Companions - Trust, Human Computer Interaction and Complex Systems* (Docotral dissertation). University of Groningen, Netherlands.

Okita, S. Y., Ng-Thow-Hing, V., & Sarvadevabhatla, R. K. (2012). Captain May I? Proxemics Study Examining Factors that Influence Distance between Humanoid Robots, Children, and Adults, during Human-Robot Interaction. In *Proceedings of the seventh annual ACM IEEE International conference on Human-Robot Interaction* (pp. 203–204).

Ososky, S., Sanders, T., Jentsch, F., Hancock, P. A., & Chen, J. Y. C. (2014). Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. *Proceedings of SPIE*, *9084*.

Parasuraman, R., & Manzey, D. H. (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. doi: 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.

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.

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 - Part A: Systems and Humans*, *30*(3), 286–297. doi: 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. doi: 10.1518/155534308X284417

Perkins, L., Miller, J. E., Hashemi, A., & Burns, G. (2010). Designing for Human-Centered Systems - Situational Risk as a Factor of Trust in Automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting Proceedings*, 2130–2134. doi: 10.1518/107118110X12829370264961

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. *6. Tagung Fahrerassistenz. Der Weg zum automatischen Fahren*.

Popken, A. (2009). *Drivers' reliance on lane keeping assistance systems as a function of the level of assistance* (Doctoral dissertation). University of Technology Chemnitz, Germany.

Pradhan, J., Ranjan, R., & Samal, A. (2015). Reaction time as a predictor of trust: Empirical evidence. *The Journal of Community Health Management*, *2*(1), 43–46.

Rajaonah, B., Anceaux, F., & Vienne, F. (2006). Study of driver trust during cooperation with adaptive cruise control. *Le travail humain*, *69*(2), 99–127. doi: 10.3917/th.692.0099

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)*, *13*(3), 257–266. doi: 10.1109/TSMC.1983.6313160

Rauch, N., Gradenegger, B., & Krüger, H.-P. (2007). Das Konzept des Situationsbewusstseins und seine Implikationen für die Fahrsicherheit. *FAT-Schriftenreihe*, *210*.

Reason, J. (1990). *Human Error*. Cambridge, MA: Cambridge University Press.

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

Richards, D., & Stedmon, A. (2016). To delegate or not to delegate: A review of control frameworks for autonomous cars. *Applied Ergonomics*, *53*, 383–388. doi: 10.1016/j.apergo.2015.10.011

Riley, V. A. (1994). *Human use of automation* (Doctoral dissertation). University of Minnesota.

Riley, V. A. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 19–35). Hillsdale, N.J.: Lawrence Erlbaum.

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*(4), 651–665.

Rouse, W. B., Cannon-Bowers, J. A., & Salas, E. (1992). The role of mental models in team performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, *22*(6), 1296–1308.

Rouse, W. B., & Morris, N. M. (1985). *On looking into the black box: prospects and limits in the search for mental models* (Tech. Rep.). Atlanta, GA: Georgia Institute of Technology.

Rovira, E., Cross, A., Leitch, E., & Bonaceto, C. (2014). Displaying Contextual Information Reduces the Costs of Imperfect Decision Automation in Rapid Retasking of ISR Assets. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*(6), 1036–1049. doi: 10.1177/0018720813519675

Ruff, H. A., Narayanan, S., & Draper, M. H. (2002). Human Interaction with Levels of Automation and Decision-Aid Fidelity in the Supervisory Control of Multiple Simulated Unmanned Air Vehicles. *Presence*, *11*(4), 335–351.

SAE International. (2014). *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems: J3016* (Tech. Rep.). Warrendale, PA: Society of Automotive Engineers.

Safar, J. A., & Turner, C. W. (2005). Validation of a Two Factor Structure for System Trust. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*(3), 497–501. doi: 10.1177/154193120504900360

Sanchez, J., Rogers, W. A., Fisk, A. D., & Rovira, E. (2011). Understanding reliance on automation: effects of error type, error distribution, age and experience. *Theoretical Issues in Ergonomics Science*, *15*(2), 134–160. doi: 10.1080/1463922X.2011.611269

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. doi: 10.1518/001872095779049516

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of Human Factors & Ergonomics* (pp. 1926–1943). New York: Wiley.

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology* (pp. 35–37). Windsor, UK: NFER Nelson.

Schwarzer, R., & Jerusalem, M. (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen: Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen.* Berlin: Freie Universität Berlin.

Seong, Y., & Bisantz, A. M. (1998). Modeling Human Trust in Complex Automated Systems. In *Studies and Analyses of Aided Adversarial Decision Making* (pp. 94–100).

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

Shen, S., & Neyens, D. M. (2014). Assessing drivers' performance when automated driver support systems fail with different levels of automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 2068–2072. doi: 10.1177/1541931214581435

Sheridan, T. B. (1988). Trustworthiness of Command and Control Systems. *IFAC Proceedings Volumes*, *21*(5), 427–431. doi: 10.1016/S1474-6670(17)53945-2

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

Sheridan, T. B. (2002). *Humans and automation: System design and research issues* (Vol. 3). Santa Monica, CA: Human Factors and Ergonomics Society.

Sheridan, T. B., & Hennessy, R. T. (1984). *Research and Modeling of Supervisory Control Behavior*. Washington, DC: National Academy Press.

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

Sheridan, T. B., & Verplank, W. L. (1978). *Human and Computer Control of Undersea Teleoperators* (Tech. Rep.). Cambridge, MA: Massachusetts Institute of Technology.

Simpson, J. A., & Brander, G. N. (1995). Seaworthy Trust: Confidence in Automated Data Fusion. In *The Human-Electronic Crew: Can We Trust the Team?* (pp. 77–82). Cambridge, UK: Defence Research Agency - Centre for Human Sciences.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-Induced "Complacency": Development of the Complacency-Potential Rating Scale. *International Journal of Aviation Psychology*, *3*(2), 111–122.

Sitkin, S. B., & Pablo, A. L. (1992). Reconceptualizing the Determinants of Risk Behavior. *The Academy of Management Review*, *17*(1), 9–38.

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, *51*, 991–1006.

Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A., & Leimeister, J. M. (2012). Understanding the Formation of Trust in IT Artifacts. *Proceedings of the International Conference on Information Systems (ICIS)*.

Söllner, M., & Leimeister, J. M. (2011). Vertrauensunterstützung für ubiquitäre Systeme: Von der Vertrauenstheorie zu Anforderungen an ubiquitäre Systeme. *Berliner Werkstatt Mensch-Maschine-Systeme*.

Söllner, M., & Leimeister, J. M. (2013). Vertrauenswürdige Gestaltung von ubiquitären Systemen: Potentiale des Einsatzes von Methoden des NeuroIS. *Berliner Werkstatt Mensch-Maschine-Systeme*.

Spain, R. D., Bustamante, E. A., & Bliss, J. P. (2008). Towards an empirically developed scale for system trust: Take two. *Human Factors and Ergonomics Society Annual Meeting Proceedings*.

Spain, R. D., & Madhavan, P. (2009). The Role of Automation Etiquette and Pedigree in Trust and Dependence. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *53*(4), 339–343.

Statista. (2015). *Autonome Fahrzeuge: Statista-Dossier*. Hamburg.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, *47*(2), 143–148. doi: 10.1016/0001-6918(81)90005-6

Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., ... Mahoney, P. (2006). Stanley: The robot that won the DARPA Grand Challenge. *Journal of Field Robotics*, *23*(9), 661–692. doi: 10.1002/rob.20147

United Nations Economic and Social Council. (1968). *Convention on Road Traffic*.

United Nations Economic and Social Council. (2014). *Report of the 68th session of the Working Party on Road Traffic Safety*.

Urmson, C., Anhalt, J., Bagnell, D., Baker, C., Bittner, R., Clark, M. N., ... Ferguson, D. (2008). Autonomous driving in urban environments: Boss and the Urban Challenge. *Journal of Field Robotics*, *25*(8), 425–466. doi: 10.1002/rob.20255

Verband der Automobilindustrie e. V. (2015). Automatisierung: Von Fahrerassistenzsystemen zum automatisierten Fahren. *VDA Magazin*.

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. doi: 10.1177/0018720812443825

VIRES Simulationstechnologie GmbH. (2014). *Virtual Test Drive*. Retrieved 22.04.2017, from `http://www.vires.com/`

Walker, G. H., Stanton, N. A., & Young, M. S. (2001). Where Is Computing Driving Cars? *International Journal of Human-Computer Interaction*, *13*(2), 203–229.

Walliser, J. C. (2011). *Trust in automated systems: the effect of automation level on trust calibration* (Master's thesis). Naval Postgraduate School, Monterey, California.

Wang, M. (2010). *Trust, situation awareness and automation use: exploring the effect of visual information degradations on human perception and performance in human-telerobot* (Doctoral dissertation). State University, Iowa.

Waytz, A., Heafner, J., & Epley, N. (2014). The Mind in the Machine: Anthropomorphism Increases Trust in an Autonomous Vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117. doi: 10.1016/j.jesp.2014.01.005

Wei, Z.-G., Macwan, A. P., & Wieringa, P. A. (1998). A Quantitative Measure for Degree of

Automation and Its Relation to System Performance and Mental Load. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *40*(2), 277–295. doi: 10.1518/001872098779480406

Wickens, C. D., Hollands, J. D., Banbury, S., & Parasuraman, R. (2016). *Engineering Psychology and Human Performance* (4th ed.). New York: Taylor & Francis.

Wiczorek, R. (2011). Entwicklung und Evaluation eines mehrdimensionalen Fragebogens zur Messung von Vertrauen in technische Systeme. In S. Schmid, M. Elepfandt, J. Adenauer, & A. Lichtenstein (Eds.), *Reflexionen und Visionen der Mensch-Maschine-Interaktion* (Vol. 33, pp. 621–626). Berlin: VDI-Verlag GmbH.

Wiener, E. L. (1981). Complacency: Is the term useful for air safety? In *Proceedings of the 26th Corporate Aviation Safety Seminar.* Denver, CO: Flight Safety Foundation.

Ye, L. R., & Johnson, P. E. (1995). The Impact of Explanation Facilities on User Acceptance of Expert Systems Advice. *MIS Quarterly*, *19*(2), 157–172. doi: 10.2307/249686

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482. doi: 10.1002/cne.920180503

Young, M. S., & Stanton, N. A. (1997). Automotive automation: Investigating the impact on driver mental workload. *International Journal of Cognitive Ergonomics*, *1*(4), 325–336.

Zeeb, K., & Schaub, M. (2014). Re- vs. Neuorientierung: Situationsgerechtes Blickverhalten beim hochautomatisierten Fahren. *Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel (AAET)*, *15*, 70–80.

Ziegler, M. (2013). Vom Assistenten zum Autopilot. *Automobil Industrie*, *3*, 22–24.

Zuboff, S. (1988). *In the age of the smart machine: The future of work and power*. New York: Basic Books.

A Appendices

A.1 Appendix A: Instructions and questionnaires

A.1.1 Instructions

Instructions Study 1

Vielen Dank für Ihre Bereitschaft zur Teilnahme an diesem Versuch. Wir werden heute eine Testfahrt zum Thema "hochautomatisiertes Fahren" durchführen. Wir werden gleich mit diesem Versuchsträger eine Autobahnfahrt absolvieren, bei der Sie von einem hochautomatisierten System unterstützt werden. Dieser Versuchsteil wird ca. 1,5 Stunden dauern. Dabei wird ausschließlich das System getestet und nicht Sie als Person. Es werden bei diesem Versuch Ihre Blickbewegungen, Fahrdaten sowie Videos aufgezeichnet. Ihre Daten werden selbstverständlich vertraulich behandelt und anonymisiert ausgewertet werden. Wenn Sie noch Fragen haben, wenden Sie sich bitte an die Versuchsleitung.
[Einsteigen]
Sie dürfen nun im Versuchsträger Platz nehmen. Bitte stellen Sie sich den Sitz und die Spiegel so ein, dass sie das Fahrzeug sicher führen können. Sie werden nun erst einmal Fragen zu Ihrer Person beantworten. Entscheiden Sie dabei möglichst spontan. Es ist wichtig, dass Sie nicht lange über die Antwort nachdenken, damit Ihre unmittelbare Einschätzung zum Tragen kommt. Es gibt keine richtige oder falsche Antwort.
[Fragebogen]
Wir werden nun mit der Testfahrt mit dem hochautomatisierten System starten. Zuvor möchte ich Ihnen dieses System und die Anzeige- und Bedienkonzepte genauer erklären. Bei Unklarheiten zur Funktionsweise können Sie mich jederzeit fragen. Das hochautomatisierte System, das Sie gleich kennen lernen werden, übernimmt auf der Autobahn die gesamte Fahraufgabe für Sie. Das heißt, das System regelt sowohl die Geschwindigkeit, als auch den Abstand zum Vorderfahrzeug und die Spurhaltung und übernimmt gegebenenfalls Manöver wie Fahrstreifenwechsel. Das hochautomatische System funktioniert in einem Geschwindigkeitsbereich zwischen 0 und 130 km/h. Während das hochautomatisierte System aktiv ist, dreht sich das Lenkrad bei Lenkbewegungen mit. Sie sind bei aktiviertem System nicht mehr dazu verpflichtet, das System und dessen Fahrzeugführung zu überwachen.
Der aktuelle Systemstatus lässt sich jederzeit im Kombi-Instrument und auf dem zentralen Statusindikator in der Mittelkonsole ablesen. Zusätzlich gibt es die vollflächige LED-Leiste in der Scheibenwurzel. Ist das hochautomatisierte System ausgeschaltet und zum gegenwärtigen Zeitpunkt nicht verfügbar, weil die Voraussetzungen zur Nutzung nicht erfüllt sind, so ist das Systemsymbol grau dargestellt (nicht auf der Autobahn). Erkennt das System eine Situation, die sich im zugelassenen Bereich bewegt, wird Ihnen die Aktivierung des Systems angeboten. Das Systemsymbol im Kombi-Instrument wird dann weiß, die LED-Leiste färbt sich türkis ein und ein Hinweiston ertönt. Sie können das System dann über ein gleichzeitiges Drücken der beiden Lenkrad-Tasten aktivieren. Bitte tun sie dies, wenn diese Situation im Versuch eintritt, denn nur durch eine entsprechend durchgehende Nutzung des Systems können Sie genug Erfahrung sammeln, um uns anschließend eine fundierte Bewertung des Systems zu geben. Nach der

Aktivierung des Systems können Sie die Hände vom Lenkrad und die Füße von den Pedalen nehmen.
Das System gibt Ihnen im Kombi-Instrument und im zentralen Statusindikator Informationen über den Systemstatus. Darüber hinaus können Sie weitere Informationen im großen Display in der Mittelkonsole sehen. Dort werden Sie eine Umfeld-Darstellung bekommen, die Ihnen eine Überprüfung der Sensorerkennung ermöglicht ("was das Fahrzeug sieht"). Der Ort dieser Umgebungsanzeige ist noch nicht optimal – die Anzeige soll zukünftig im Kombi-Instrument dargestellt werden. Heute soll es daher eher um die Inhalte der Anzeige gehen – schauen Sie sich die Anzeige daher bitte genau an. Bitte achten Sie darauf, welche Informationen Ihnen dort vom System zur Verfügung gestellt werden, um ein vollständiges Verständnis über das System zu erreichen.
Sie können das System jederzeit mit einer Bremsung oder Betätigung der beiden Lenkradtasten übersteuern. Diesmal wird das System die Steuerung außerdem sofort an Sie übergeben, wenn Sie nach dem Lenkrad greifen oder ein Lenkmanöver vornehmen. Danach müssen Sie die Fahrzeugführung sofort manuell übernehmen. Um das System wieder zu aktivieren, betätigen Sie bitte erneut die Lenkradtasten, wenn das System verfügbar ist.
Der Versuchsablauf sieht vor, dass Sie manuell auf die Autobahn auffahren, da das System nur auf der Autobahn aktiv regeln kann. Das hochautomatisierte System kann also nur auf der geraden Autobahn aktiviert werden. Dies tun Sie nach der Verfügbarkeitsanzeige über die beiden Lenkradtasten. Eine Übernahme am Ende jeder Teilstrecke wird Ihnen akustisch und in den Displays angezeigt. Bitte übernehmen Sie in diesen Phasen die Fahrzeugführung und fahren an den entsprechenden Stellen manuell von der Autobahn ab.
Im folgenden Abschnitt werden zur Eingewöhnung zunächst keine Fahrstreifenwechsel vom Fahrzeug durchgeführt werden. Sollten Sie den Wunsch haben, den Fahrstreifen dennoch zu wechseln, können Sie das System deaktivieren und selbst auf einen anderen Fahrstreifen wechseln. Anschließend können Sie das System wieder aktivieren. Haben Sie dazu noch Fragen?
Während der Fahrt werden wir Ihre Blickbewegung mit Hilfe einer Augenkamera (zeichnet nur Pupille auf) und einer Umfeldkamera (zeichnet Sichtfeld auf) messen, die Sie ähnlich einer Brille tragen. Aus diesem Grund sollten Sie sich möglichst ruhig verhalten und sich möglichst nicht mit der Hand ins Gesicht oder an den Kopf fassen. Beides stört die Erfassung der Blickrichtung. Die Blickdaten sind wichtig, um später Fahrdaten zu interpretieren. Bitte setzen Sie sich die Brille nun auf. Wir werden nun mit einer kurzen Kalibrierung beginnen.
[Kalibrierung Eye-Tracking]
Wenn Sie nun keine weiteren Fragen haben, folgt die eigentliche Versuchsfahrt.
[Versuchsfahrt mit Befragungen]
Bitte fahren Sie hier ab und suchen Sie sich einen Parkplatz. Bitte machen Sie den Motor nicht aus. Sie haben nun den letzten Streckenabschnitt mit dem hochautomatisierten System erlebt. Es erfolgt nun noch eine Abschlussbefragung. Auch hier entscheiden Sie bitte möglichst spontan. Es ist wichtig, dass Sie nicht lange über die Antwort nachdenken, damit Ihre unmittelbare Einschätzung zum Tragen kommt. Es gibt keine richtige oder falsche Antwort.
Vielen Dank für Ihre Teilnahme.

Instructions Study 2

Vielen Dank für Ihre Bereitschaft zur Teilnahme an diesem Versuch. Wir werden heute eine Simulator-Fahrt zum Thema "hochautomatisiertes Fahren" durchführen. Sie werden gleich in einer Sitzkiste mehrere Autobahnfahrten absolvieren, bei denen Sie von einem hochautomatisierten System unterstützt werden. Dieser Versuch wird insgesamt ca. 2 Stunden dauern. Dabei wird ausschließlich das System getestet und nicht Sie als Person. Es werden bei diesem Versuch Ihre Blickbewegungen sowie Fahrdaten aufgezeichnet. Ihre Daten werden selbstverständlich vertraulich behandelt und anonymisiert ausgewertet werden. Wenn Sie noch Fragen haben, wenden Sie sich bitte an die Versuchsleitung. Sie werden nun erst einmal Fragen zu Ihrer Person beantworten. Entscheiden Sie dabei möglichst spontan. Es ist wichtig, dass Sie nicht lange über die Antwort nachdenken, damit Ihre unmittelbare Einschätzung zum Tragen kommt. Es gibt keine „richtige" oder „falsche" Antwort.
[Fragebogen]
Sie dürfen nun in der Sitzkiste Platz nehmen. Bitte stellen Sie sich den Sitz bitte so ein, dass sie das Fahrzeug sicher führen können. Die Spiegel werden von der Versuchsleitung für Sie passend eingestellt.
[Einsteigen]
Während der Fahrt werden wir Ihre Blickbewegung mit Hilfe einer Augen-Kamera (zeichnet nur Pupille auf) und einer Umfeld-Kamera (zeichnet Sichtfeld auf) messen, die Sie ähnlich einer Brille tragen. Aus diesem Grund sollten Sie sich möglichst ruhig verhalten und sich möglichst nicht mit der Hand ins Gesicht oder an den Kopf fassen. Beides stört die Erfassung der Blickrichtung. Die Blickdaten sind wichtig, um später Fahrdaten zu interpretieren. Bitte setzen Sie sich die Brille nun auf. Wir werden mit einer kurzen Kalibrierung beginnen.
[Kalibrierung Eye-Tracking]
Wir werden nun mit einer Übungsfahrt mit dem hochautomatisierten System starten.

- *Kontrollgruppe: Kurze Einweisung.* Wenn das System verfügbar ist, können Sie die beiden Tasten auf dem Lenkrad drücken, um das System zu aktivieren. Das hochautomatisierte System übernimmt dann auf der Autobahn die gesamte Fahraufgabe für Sie. Sie sind bei aktiviertem System nicht mehr dazu verpflichtet, das System und dessen Fahrzeugführung zu überwachen. Eine Übernahme am Ende jeder Teilstrecke wird Ihnen akustisch und in den Displays angezeigt. Bitte übernehmen Sie in diesen Phasen die Fahrzeugführung und fahren an den entsprechenden Stellen manuell von der Autobahn ab.
- *Versuchsgruppe: Detaillierte Einweisung.* Bei der Übungsfahrt werde ich Ihnen das System und die Anzeige- und Bedienkonzepte genauer erklären. Bei Unklarheiten zur Funktionsweise können Sie mich jederzeit fragen. Das hochautomatisierte System, das Sie nun kennen lernen, übernimmt auf der Autobahn die gesamte Fahraufgabe für Sie. Das heißt, das System regelt sowohl die Geschwindigkeit, als auch den Abstand zum Vorderfahrzeug und die Spurhaltung und übernimmt gegebenenfalls Manöver wie Fahrstreifenwechsel. Das hochautomatische System funktioniert in einem Geschwindigkeitsbereich zwischen 0 und 130 km/h. Während das hochautomatisierte System aktiv ist, dreht sich das Lenkrad bei Lenkbewegungen mit. Sie sind bei aktiviertem System nicht mehr dazu verpflichtet, das System und dessen Fahrzeugführung zu überwachen. Der aktuelle Systemstatus lässt sich jederzeit im Kombi-Instrument anhand des Auto-Pilot Symbols ablesen (runder Statusindikator in der linken Tube). Ist das hochautomatisierte System ausgeschaltet und zum gegenwärtigen Zeitpunkt nicht verfügbar, weil die Voraussetzungen zur Nutzung nicht erfüllt sind (beispielsweise nicht auf der Autobahn), so wird in der linken Tube der Abstand bis zu einem verfügbaren Streckenabschnitt eingeblendet. Erkennt das System eine Situation, die sich im zugelassenen Bereich bewegt, wird Ihnen die Aktivierung des Systems angeboten. Die vollflächige LED-Leiste in der Scheibenwurzel färbt sich dann oberhalb des Kombi-Instrumentes weiß ein, die Sprachausgabe „Autobahnpilot verfügbar" ertönt, und im Kombi-Instrument erscheint ein Pop-up, welches Sie auffordert, die beiden Tasten auf dem Lenkrad zu

drücken. Sie können das System dann über ein gleichzeitiges Drücken der beiden Lenkrad-Tasten aktivieren. Bitte tun sie dies, wenn diese Situation im Versuch eintritt, denn nur durch eine entsprechend durchgehende Nutzung des Systems können Sie genug Erfahrung sammeln, um uns anschließend eine fundierte Bewertung des Systems zu geben. Bitte aktivieren Sie das System immer auf dem rechten Fahrstreifen.
Der Versuchsablauf sieht vor, dass Sie manuell auf die Autobahn auffahren, da das System nur auf der Autobahn aktiv regeln kann. Das hochautomatisierte System kann also nur auf der geraden Autobahn aktiviert werden. Dies tun Sie nach der Verfügbarkeitsanzeige über die beiden Lenkradtasten und wenn Sie sich in der Mitte eines Fahrstreifens befinden. Nach der Aktivierung des Systems können Sie die Hände vom Lenkrad und die Füße von den Pedalen nehmen. Die türkise Farbe der LED-Leiste gibt Ihnen dauerhaft den Systemstatus des Auto-Piloten an.

- *HMI-Gruppe A.* Im Kombi-Instrument gibt Ihnen das System ebenfalls Informationen über den Systemstatus.
- *HMI-Gruppe B.* Im Kombi-Instrument gibt Ihnen das System weitere Informationen über den Systemstatus. Dort werden Sie eine Umfeld-Darstellung bekommen, die Ihnen eine Überprüfung der Sensorerkennung ermöglicht („was das Fahrzeug sieht“). Die Verortung des Fahrzeugs auf der Straße wird dargestellt, sowie die beiden rechts und links vom Fahrzeug befindlichen Fahrstreifen. Kleine Fahrzeuge im Display zeigen an, dass das System den umgebenden Verkehr erkannt hat.
- *HMI-Gruppe C.* Im Kombi-Instrument gibt Ihnen das System weitere Informationen über den Systemstatus. Dort werden Sie eine Umfeld-Darstellung bekommen, die Ihnen eine Überprüfung der Sensorerkennung ermöglicht („was das Fahrzeug sieht“). Die Verortung des Fahrzeugs auf der Straße wird dargestellt, sowie die beiden rechts und links vom Fahrzeug befindlichen Fahrstreifen. Kleine Fahrzeuge im Display zeigen an, dass das System den umgebenden Verkehr erkannt hat. Darüber hinaus werden Ihnen Manöver des Fahrzeugs ebenfalls im Kombi-Instrument angezeigt. Zusätzlich zu einem Pfeil, der Ihnen die Richtung des geplanten Manövers anzeigt, wird auch der Grund des Manövers abgebildet (zum Beispiel “Überholmanöver” oder “Rechtsfahrgebot”).

Heute soll es um die Inhalte dieser Anzeige gehen – schauen Sie sich die Anzeige daher bitte genau an. Bitte achten Sie darauf, welche Informationen Ihnen dort vom System zur Verfügung gestellt werden, um ein vollständiges Verständnis über das System zu erreichen.
Eine Übernahme am Ende jeder Teilstrecke wird Ihnen akustisch und in den Displays angezeigt. Bitte übernehmen Sie in diesen Phasen die Fahrzeugführung und fahren an den entsprechenden Stellen manuell von der Autobahn ab. Sie können das System theoretisch auch jederzeit mit einer Bremsung oder Betätigung der beiden Lenkradtasten übersteuern, wenn Sie sich einmal unwohl fühlen sollten. Das System wird die Steuerung auch sofort an Sie übergeben, wenn Sie ein Lenkmanöver vornehmen. Danach müssen Sie die Fahrzeugführung sofort manuell übernehmen. Um das System wieder zu aktivieren, betätigen Sie bitte erneut die Lenkradtasten, wenn das System verfügbar ist (bitte ausschließlich auf dem rechten Fahrstreifen aktivieren).

Haben Sie dazu noch Fragen? Wenn Sie nun keine weiteren Fragen haben, folgt die eigentliche Versuchsfahrt.
[Versuchsfahrten mit Befragungen]
Es erfolgt nun noch eine Abschlussbefragung. Auch hier entscheiden Sie bitte möglichst spontan. Es ist wichtig, dass Sie nicht lange über die Antwort nachdenken, damit Ihre unmittelbare Einschätzung zum Tragen kommt. Es gibt keine richtige oder falsche Antwort.
Vielen Dank für Ihre Teilnahme.

Instructions Study 3

Vielen Dank für Ihre Bereitschaft zur Teilnahme an diesem Versuch. Wir werden heute (und bei den folgenden Terminen) eine Testfahrt zum Thema "hochautomatisiertes Fahren" durchführen. Wir werden gleich mit diesem Versuchsträger eine Autobahnfahrt absolvieren, bei der Sie von einem hochautomatisierten System unterstützt werden. Uns interessiert, wie Sie sich an den Umgang mit dem System gewöhnen. Neben der Versuchsleitung wird eine zweite Person den Versuch begleiten, die sich um die Datenaufzeichnung und die Technik im Hintergrund kümmert (aber ansonsten nicht in den Versuch involviert ist). Dieser Versuchsteil wird ca. 2 Stunden dauern. Dabei wird ausschließlich das System getestet und nicht Sie als Person. Es werden bei diesem Versuch Ihre Blickbewegungen, Fahrdaten sowie Videos aufgezeichnet. Ihre Daten werden selbstverständlich vertraulich behandelt und anonymisiert ausgewertet werden. Haben Sie noch Fragen? Sie dürfen nun im Fahrzeug Platz nehmen. Bitte stellen Sie sich den Sitz und die Spiegel bitte so ein, dass sie das Fahrzeug sicher führen können.
[Einsteigen]
Sie werden nun erst einmal Fragen zu Ihrer Person beantworten. Entscheiden Sie dabei möglichst spontan. Es ist wichtig, dass Sie nicht lange über die Antwort nachdenken, damit Ihre unmittelbare Einschätzung zum Tragen kommt. Es gibt keine richtige oder falsche Antwort.
[Fragebogen]
Wir werden nun mit der Testfahrt mit dem hochautomatisierten System starten. Zuvor möchte ich Ihnen dieses System und die Anzeige- und Bedienkonzepte genauer erklären. Bei Unklarheiten zur Funktionsweise können Sie mich jederzeit fragen. Sie können gleich in der Mittelkonsole die Navigation starten (mithilfe des großen Drehdrückstellers) – wir möchten heute auf der Autobahn nach Magdeburg fahren. Die Navigation wird Ihnen den Weg dorthin zeigen.
Der Versuchsablauf sieht vor, dass Sie manuell auf die Autobahn auffahren. Das hochautomatisierte System, das Sie gleich kennen lernen werden, übernimmt auf der Autobahn die gesamte Fahraufgabe für Sie. Das heißt, das System regelt sowohl die Geschwindigkeit, als auch den Abstand zum Vorderfahrzeug und die Spurhaltung und übernimmt gegebenenfalls Manöver wie Fahrstreifenwechsel. Das hochautomatische System funktioniert in einem Geschwindigkeitsbereich zwischen 0 und 130 km/h. Während das hochautomatisierte System aktiv ist, dreht sich das Lenkrad bei Lenkbewegungen mit. Sie sind bei aktiviertem System nicht mehr dazu verpflichtet, das System und dessen Fahrzeugführung zu überwachen. Ist das hochautomatisierte System ausgeschaltet und zum gegenwärtigen Zeitpunkt nicht verfügbar, weil die Voraussetzungen zur Nutzung nicht erfüllt sind, so ist das Systemsymbol grau dargestellt (bei Nachfrage: nicht auf der Autobahn). Erkennt das System eine Situation, die sich im zugelassenen Bereich bewegt, wird Ihnen die Aktivierung des Systems angeboten. Das Systemsymbol im Kombi-Instrument wird dann weiß, die LED-Leiste färbt sich türkis ein und ein Hinweiston ertönt. Sie können das System dann über ein gleichzeitiges Drücken der beiden Lenkrad-Tasten aktivieren (möglichst in der Mitte des Fahrstreifens). Bitte tun sie dies, wenn diese Situation im Versuch eintritt, denn nur durch eine entsprechend durchgehende Nutzung des Systems können Sie genug Erfahrung sammeln, um uns anschließend eine fundierte Bewertung des Systems zu geben. Nach der Aktivierung des Systems können Sie die Hände vom Lenkrad und die Füße von den Pedalen nehmen. Der aktuelle Systemstatus lässt sich jederzeit im Kombi-Display ablesen.

- *HMI-Gruppe A.* Dort werden Sie eine Kilometerangabe sehen, die Ihnen anzeigt, wie lange das automatisierte System noch verfügbar ist.
- *HMI-Gruppe B.* Dort werden Sie eine Kilometerangabe sehen, die Ihnen anzeigt, wie lange das automatisierte System noch verfügbar ist. Zudem werden Sie eine Umfeld-Darstellung bekommen, die Ihnen eine Überprüfung der Sensorerkennung ermöglicht ("was das Fahrzeug sieht"). Die Verortung des Fahrzeugs auf der Straße wird dargestellt, sowie die beiden rechts und links vom Fahrzeug befindlichen Fahrstreifen. Kleine

> Fahrzeuge im Display zeigen an, dass das System den umgebenden Verkehr erkannt hat. In der Anzeige werden Ihnen auch geplante Manöver (wie Spurwechsel) des Fahrzeugs mithilfe eines Pfeils und eine Begründung für das Verhalten des Fahrzeugs angezeigt. Zusätzlich gibt es eine vollflächige LED-Leiste in der Scheibenwurzel.

Eine Übernahme am Ende jeder Teilstrecke wird Ihnen akustisch und in den Displays angezeigt. Bitte übernehmen Sie in diesen Phasen die Fahrzeugführung und fahren an den entsprechenden Stellen manuell von der Autobahn ab. Sie können das System jederzeit mit einer Bremsung oder der Betätigung der beiden Lenkradtasten übersteuern, wenn Sie sich einmal unwohl bei der Systemnutzung fühlen sollten. Das System wird die Steuerung auch sofort an Sie übergeben, wenn Sie ein Lenkmanöver vornehmen. Danach müssen Sie die Fahrzeugführung sofort manuell übernehmen. Um das System wieder zu aktivieren, betätigen Sie bitte erneut die Lenkradtasten, wenn das System verfügbar ist. Haben Sie dazu noch Fragen?

Während der Fahrt werden wir Ihre Blickbewegung mit Hilfe einer Augen-Kamera (zeichnet nur Pupille auf) und einer Umfeld-Kamera (zeichnet Sichtfeld auf) messen, die Sie ähnlich einer Brille tragen. Aus diesem Grund sollten Sie sich möglichst ruhig verhalten und sich möglichst nicht mit der Hand ins Gesicht oder an den Kopf fassen. Beides stört die Erfassung der Blickrichtung. Die Blickdaten sind wichtig, um später Fahrdaten zu interpretieren. Bitte setzen Sie sich die Brille nun auf. Wir werden mit einer kurzen Kalibrierung beginnen.

[Kalibrierung Eye-Tracking]

Sie werden nun zwei ca. 60-minütige Streckenabschnitte erleben, welchen Sie unterstützt durch ein hochautomatisiertes System durchfahren. Das System übernimmt für Sie die Fahrzeugführung und hält sich an die Straßenverkehrsordnung. Wenn Sie möchten, können Sie während der Fahrt auch etwas anderes machen: zum Beispiel sich ein Video über das Tablet ansehen. Bitte beachten Sie zeitgleich aber auch die Anzeigen des Systems, da Sie diese nach der Testfahrt bewerten sollen.

[Versuchsfahrten mit Befragungen]

Bitte fahren Sie hier ab und suchen einen Parkplatz. Es erfolgt nun noch eine Abschlussbefragung. Auch hier entscheiden Sie bitte möglichst spontan. Es ist wichtig, dass Sie nicht lange über die Antwort nachdenken, damit Ihre unmittelbare Einschätzung zum Tragen kommt. Es gibt keine richtige oder falsche Antwort.

Vielen Dank für Ihre Teilnahme.

A.1.2 Rating scales

Table A.1
5-point Likert-type rating scale used for personality and attitude questionnaires

strongly disagree	**disagree**	**neither agree nor disagree**	**agree**	**strongly agree**
trifft gar nicht zu	trifft nicht zu	teils / teils	trifft eher zu	trifft voll zu
1	2	3	4	5

Table A.2
15-point rating scale used for single items, based on Heller (1985)

	very low			**low**			**neutral**			**high**			**very high**	
	sehr gering			gering			neutral			hoch			sehr hoch	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

A.1.3 Questionnaires

Trust questionnaires

Table A.3
System Trust Scale by Jian et al. (2000)

Number	Item	Subscale
1	The system is deceptive.	Distrust
2	The system behaves in an underhanded manner.	
3	I am suspicious of the system's intent, action, or outputs.	
4	I am wary of the system.	
5	The system's actions will have a harmful or injurious outcome.	
6	I am confident in the system.	Trust
7	The system provides security.	
8	The system has integrity.	
9	The system is dependable.	
10	The system is reliable.	
11	I can trust the system.	
12	I am familiar with the system.	

Table A.4
Scale items of the Human-Computer Trust scale by Madsen and Gregor (2000)

Number	Item	Subscale
R1	The system always provides the advice I require to make my decision.	Perceived Reliability
R2	The system performs reliably.	
R3	The system responds the same way under the same conditions at different times.	
R4	I can rely on the system to function properly.	
R5	The system analyzes problems consistently.	
T1	The system uses appropriate methods to reach decisions.	Perceived Technical Competence
T2	The system has sound knowledge about this type of problem built into it.	
T3	The advice the system produces is as good as that which a highly competent person could produce.	

Scale items of the Human-Computer Trust scale by Madsen and Gregor (2000) (continued)

Number	Item	Subscale
T4	The system correctly uses the information I enter.	
T5	The system makes use of all the knowledge and information available to it to produce its solution to the problem.	
U1	I know what will happen the next time I use the system because I understand how it behaves.	Perceived Understandability
U2	I understand how the system will assist me with decisions I have to make.	
U3	Although I may not know exactly how the system works, I know how to use it to make decisions about the problem.	
U4	It is easy to follow what the system does.	
U5	I recognize what I should do to get the advice I need from the system the next time I use it.	
F1	I believe advice from the system even when I don't know for certain that it is correct.	Faith
F2	When I am uncertain about a decision I believe the system rather than myself.	
F3	If I am not sure about a decision, I have faith that the system will provide the best solution.	
F4	When the system gives unusual advice I am confident that the advice is correct.	
F5	Even if I have no reason to expect the system will be able to solve a difficult problem, I still feel certain that it will.	
P1	I would feel a sense of loss if the system was unavailable and I could no longer use it.	Personal Attachment
P2	I feel a sense of attachment to using the system.	
P3	I find the system suitable to my style of decision making.	
P4	I like using the system for decision making.	
P5	I have a personal preference for making decisions with the system.	

Table A.5

Shortened trust in automation scale by Lee and Moray (1992)

Number	**German item** (own translation)	**English item** (original)	**Subscale**
1	Wie gut konnten Sie das Verhalten des hochautomatisiert fahrenden Fahrzeugs in den eben erlebten Situationen vorhersagen?	To what extent can the system's behavior be predicted from moment to moment?	Predictability
2	Wie sehr konnten Sie sich in den eben erlebten Situationen darauf verlassen, dass das hochautomatisiert fahrende System funktioniert?	To what extent can you count on the system to do its job?	Dependability
3	Wie hoch ist Ihr Glaube daran, dass das hochautomatisiert fahrende Fahrzeug mit Situationen dieser Art jederzeit umgehen kann?	What degree of faith do you have that the system will be able to cope with all situations in the future?	Faith
4	Wie hoch ist Ihr Vertrauen in das hochautomatisiert fahrende System nach der eben erlebten Fahrt?	Overall how much do you trust the system?	Overall trust

Demographics & driving behavior

Table A.6

Demographic questions used in the main user studies (German translation in parentheses)

Question	**Answer format**
Gender (Geschlecht)	☐ male (männlich) ☐ female (weiblich)
Age (Alter)	___ years (Jahre)
Years since driving license (Führerscheinbesitz)	___ years (Jahre)

Table A.7
Questions regarding driving behavior used in the main user studies (German translation in in parentheses)

Question	Answer format
How often do you drive? (Wie oft fahren Sie Auto?)	☐ Never (Nie) ☐ Rarely (Selten) ☐ At least once a month (Mehr als einmal im Monat) ☐ At least once a week (Mehr als einmal in der Woche) ☐ Almost daily (Fast täglich)
Do you prefer to drive yourself or let someone else drive? (Fahren Sie lieber selbst oder lassen Sie jemand anderes fahren?)	☐ Prefer to drive (Ich fahre lieber selbst) ☐ Let someone else drive (Ich lasse lieber jemand anderes fahren)
How would you describe your driving style? (Wie würden Sie Ihren eigenen Fahrstil einschätzen?)	☐ very defensive (sehr defensiv) ☐ somewhat defensive (eher defensiv) ☐ somewhat aggressive (eher sportlich) ☐ very aggressive (sehr sportlich)
How well does each of the following phrases describe your opinion? (Wie gut spiegeln die folgenden Aussagen Ihre Meinung wider?) - I am a better driver compared with the average (Ich bin ein besserer Autofahrer im Vergleich zum allgemeinen Durchschnitt) - I have less risk in traffic compared with the average (Mein Unfall- und Gefahren-Risiko während des Autofahrens ist geringer als der Durchschnitt) - I am better in coping with hazards in traffic compared with the average (Ich kann besser mit Gefahrensituationen im Straßenverkehr umgehen als der Durchschnitt)	☐ strongly disagree (trifft gar nicht zu) ☐ disagree (trifft nicht zu) ☐ neither agree nor disagree (teils / teils) ☐ agree (trifft eher zu) ☐ strongly agree (trifft voll zu)
Please state how experienced you are concerning the use of the following driver assistance systems. (Haben Sie Erfahrung mit den folgenden Fahrerassistenzsystemen?) – CC: Cruise Control – ACC: Adaptive Cruise Control – HC: Heading Control	☐ No experience (Keine Erfahrung) ☐ Little experience (Wenig Erfahrung) ☐ Used to the system (Gewöhnt an System) ☐ Permanently using the system (Ständige Nutzung)

Rating of the test drive

Table A.8
Questions regarding driver's state, perceived system performance and system transparency

German item	English translation
Wie nervös haben Sie sich in den eben erlebten Fahrsituationen gefühlt?	How stressed did you feel during the drive?
Wie würden Sie die eben erlebte Fahrleistung des hochautomatisiert fahrenden Fahrzeugs bewerten?	How would you rate the vehicle's driving performance during the drive?
Wie nützlich empfinden Sie die Anzeige während der hochautomatisierten Fahrt?	How would you rate the usefulness of the interface during the drive?

Personality and attitude questionnaires

Table A.9
Extraversion questionnaire (short form of the NEO PI-R IPIP representation, Johnson, 2006)

Number	German item	English item
1	Auf Partys spreche ich mit vielen verschiedenen Leuten.	I talk to a lot of different people at parties.
2	Ich freunde mich schnell mit Leuten an.	I make friends easily.
3	Ich ziehe es vor, allein zu sein.	I prefer to be alone.
4	Ich vermeide den Umgang mit anderen.	I avoid contact with others.
5	Ich sehe das Leben von seiner Schokoladenseite.	I look at the bright side of life.
6	Ich habe viel Spaß.	I have a lot of fun.
7	Ich fühle mich in Gesellschaft anderer wohl.	I feel comfortable around others.
8	Ich liebe das Leben.	I love life.
9	Ich halte andere auf Distanz.	I keep others at a distance.
10	Ich strahle Freude aus.	I radiate joy.
11	Ich vermeide Menschenmengen.	I avoid crowds.
12	Ich liebe große Partys.	I love large parties.

Table A.10
Desire for control scale by Burger and Cooper (1979)

Number	**German item** (own translation)	**English item** (original)	**Subscale**
1	Ich würde lieber eine leitende Rolle einnehmen als eine nachfolgende Rolle.	I would prefer to be a leader rather than a follower.	Desire for leadership and independence (control others)
2	Ich mag es lieber, wenn jemand anderes die Führungsrolle in einem Gruppenprojekt übernimmt.	I would rather someone else took over the leadership role when I'm involved in a group project.	
3	Ich sehe mich selbst eher in der Lage als andere, bestimmte Situationen zu bewältigen.	I consider myself to be generally more capable of handling situations than others are.	
4	Es gibt viele Situationen in denen ich es bevorzugen würde, nur eine einzige Wahl zu haben anstatt mich zwischen mehreren Alternativen entscheiden zu müssen.	There are many situations in which I would prefer only one choice rather than having to make a decision.	
5	Ich warte lieber ab und hoffe, dass eine andere Person das Problem löst, so dass ich mich nicht damit auseinandersetzen muss.	I like to wait and see if someone else is going to solve a problem so that I don't have to be bothered by it.	
6	Ich bevorzuge einen Beruf, in dem ich viel Kontrolle darüber habe, was ich tue und wann ich es tue.	I prefer a job where I have a lot of control over what I do and when I do it.	
7	Andere wissen in der Regel, was das Beste für mich ist.	Others usually know what is best for me.	
8	Ich überprüfe alles an einem Auto sorgfältig, bevor ich zu einer längeren Fahrt aufbreche.	I am careful to check everything on an automobile before I leave for a long trip.	Desire for not having to take decisions (relinquish control)
9	Ich mag es, meine eigenen Entscheidungen zu treffen.	I enjoy making my own decisions.	
10	Wenn ich Auto fahre, versuche ich Situationen zu vermeiden, in denen ich durch den Fehler einer anderen Person verletzt werden könnte.	When driving, I try to avoid putting myself in a situation where I could be hurt by someone else's mistake.	
11	Ich bevorzuge es, Situationen zu vermeiden in denen eine andere Person mir sagen muss, was ich tun sollte.	I prefer to avoid situations where someone else has to tell me what it is I should be doing.	

Desire for control scale by Burger and Cooper (1979) (continued)

Number	**German item** (own translation)	**English item** (original)	**Subscale**
12	Ich engagiere mich gern politisch, weil ich so viel Einfluss wie möglich auf die Regierung haben möchte.	I enjoy political participation because I want to have as much of a say in running government as possible.	Desire for determining own life (control self)
13	Ich mag es, andere Personen in ihrem Handeln beeinflussen zu können.	I enjoy being able to influence the actions of others.	
14	Ich würde lieber mein eigenes Unternehmen leiten und meine eigenen Fehler machen, als auf die Anweisungen einer anderen Person zu hören.	I'd rather run my own business and make my own mistakes than listen to someone else's orders.	
15	Ich gebe lieber Anweisungen als sie zu bekommen.	When it comes to orders, I would rather give them than receive them.	
16	Wenn ich ein Problem erkenne, bevorzuge ich, etwas zu tun anstatt nur daneben zu sitzen.	When I see a problem I prefer to do something about it rather than sit by and let it continue.	
17	Ich wünsche mir, viele der alltäglichen Entscheidungen einer anderen Person überlassen zu können.	I wish I could push many of life's daily decisions off on someone else.	
18	Ich versuche Situationen zu vermeiden, in denen eine andere Person mir sagt was zu tun ist.	I try to avoid situations where someone else tells me what to do.	
19	Ich mag es, Kontrolle über mein eigenes Schicksal zu haben.	I enjoy having control over my own destiny.	
20	Ich verschaffe mir gerne einen umfassenden Überblick über eine Aufgabe, bevor ich anfange.	I like to get a good idea of what a job is all about before I begin.	

Table A.11
Risk taking behavior questionnaire by Desai (2012)

Number	**German item** (own translation)	**English item** (original)
1	Ich teste mich gerne ab und an selbst, indem ich etwas riskiere.	I like to test myself every now and them by doing something a little risky.
2	Manchmal nehme ich nur zum Spaß ein Risiko auf mich.	Sometimes I will take a risk just for the fun of it.
3	Manchmal finde ich es spannend, Dinge zu tun für die ich in Schwierigkeiten geraten könnte.	I sometimes find it exciting to do things for which I might get into trouble.
4	Aufregung und Abenteuer sind mir wichtiger als Sicherheit.	Excitement and adventure are more important to me than security.

Table A.12
Self-efficacy questionnaire by Schwarzer and Jerusalem (1995)

Number	**German item** (Schwarzer & Jerusalem, 1999)	**English item** (Schwarzer & Jerusalem, 1995)
1	Wenn sich Widerstände auftun, finde ich Mittel und Wege, mich durchzusetzen.	I can always manage to solve difficult problems if I try hard enough.
2	Die Lösung schwieriger Probleme gelingt mir immer, wenn ich mich darum bemühe.	If someone opposes me, I can find means and ways to get what I want.
3	Es bereitet mir keine Schwierigkeiten, meine Absichten und Ziele zu verwirklichen.	It is easy for me to stick to my goals and accomplish my goals.
4	In unerwarteten Situationen weiß ich immer, wie ich mich verhalten soll.	I am confident that I could deal efficiently with unexpected events.
5	Auch bei überraschenden Ereignissen glaube ich, dass ich gut mit ihnen zurechtkommen kann.	Thanks to my resourcefulness, I can handle unforeseen situations.
6	Schwierigkeiten sehe ich gelassen entgegen, weil ich meinen Fähigkeiten immer vertrauen kann.	I can remain calm when facing difficulties because I can rely on my coping abilities.
7	Was auch immer passiert, ich werde schon klarkommen.	No matter what comes in my way, I am usually able to handle it.
8	Für jedes Problem kann ich eine Lösung finden.	If I am in trouble, I can usually think of something to do.
9	Wenn eine neue Sache auf mich zukommt, weiß ich, wie ich damit umgehen kann.	When I am confronted with a problem, I can find several solutions.
10	Wenn ein Problem auftaucht, kann ich es aus eigener Kraft meistern.	I can solve most problems if I invest the necessary effort.

Table A.13
Acceptance of technology questionnaire by Karrer et al. (2009)

Number	**German item** (original)	**English item** (own translation)	Subscale
1	Ich informiere mich über elektronische Geräte, auch wenn ich keine Kaufabsicht habe.	I inform myself about electronic devices, even if I have no intention to buy.	Begeisterung
2	Ich liebe es, neue elektronische Geräte zu besitzen.	I love to own new electronic devices.	
3	Ich bin begeistert, wenn ein neues elektronisches Gerät auf den Markt kommt.	I am thrilled when a new electronic device comes to market.	
4	Ich gehe gern in den Fachhandel für elektronische Geräte.	I like to go to the local dealer for electronic devices.	
5	Es macht mir Spaß, ein elektronisches Gerät auszuprobieren.	I enjoy trying out an electronic device.	
6	Ich kenne die meisten Funktionen der elektronischen Geräte, die ich besitze.	I know most of the features of the electronic devices that I own.	Kompetenz
7	Ich habe bzw. hätte Verständnisprobleme beim Lesen von Elektronik und Computerzeitschriften.	I have or would have problems understanding electronics and computer magazines.	
8	Es fällt mir leicht, die Bedienung eines elektronischen Geräts zu lernen.	It is easy for me to learn to operate an electronic device.	
9	Ich kenne mich im Bereich elektronischer Geräte aus.	I know a lot about electronic devices.	
10	Elektronische Geräte helfen, an Informationen zu gelangen.	Electronic devices help to get information.	Positive Einstellung
11	Elektronische Geräte ermöglichen einen hohen Lebensstandard.	Electronic devices enable a high standard of living.	
12	Elektronische Geräte erhöhen die Sicherheit.	Electronic devices increase security.	
13	Elektronische Geräte machen unabhängig.	Electronic devices make you independent.	
14	Elektronische Geräte erleichtern mir den Alltag.	Electronic devices make everyday life easier for me.	

Acceptance of technology questionnaire Karrer et al. (2009) (continued)

Number	**German item** (original)	**English item** (own translation)	**Subscale**
15	Elektronische Geräte verringern den persönlichen Kontakt zwischen den Menschen.	Electronic devices reduce personal contact between people.	Negative Einstellung
16	Elektronische Geräte verursachen Stress.	Electronic devices cause stress.	
17	Elektronische Geräte machen krank.	Electronic devices make you sick.	
18	Elektronische Geräte machen vieles umständlicher.	Electronic devices make things more complicated.	
19	Elektronische Geräte führen zu geistiger Verarmung.	Electronic devices lead to mental depletion.	

A.2 Appendix B: Study data

A.2.1 Overview variables study 1

Table A.14
Scale means, standard deviations, reliabilities, and intercorrelations of scales used in the first user study (N = 28).

	Scale	*M*	*SD*	1	2	3	4	5	6	7	8	9	10
1	Age	36.61	9.37	–	-.21	-.06	.27	.11	.20	.19	-.50**	-.51**	-.44*
2	Gender	1.29	0.46		–	-.31	-.36	-.18	-.21	-.47*	.01	-.13	-.15
3	Acceptance of technology	3.92	0.46			*.85*	.51**	.57**	.21	.20	.41*	.44*	.47*
4	Desire for control	3.72	0.27				*.49*	.39*	.37	.30	-.14	-.15	-.08
5	Extraversion	3.91	0.44					*.78*	.01	.21	.17	.17	.22
6	Risk-taking behavior	3.12	0.74						*.75*	.11	-.20	-.20	-.24
7	Self-efficacy	3.99	0.37							*.81*	.04	.24	.22
8	Perceived performance	11.79	2.22								–	.71***	.82**
9	Initial trust	10.74	2.97									*.94*	.95***
10	Overall trust	10.93	2.54										*.94*

Reliabilities for multi-item measures are displayed in the diagonal.
Correlations between the variables are based on an n of 28. Male coded as 1; female coded as 2.
**p* < .05; ***p* < .01; ****p* < .001

A.2.2 Overview variables study 2

Table A.15
Scale means, standard deviations, reliabilities, and intercorrelations of scales used in the second user study (N = 72).

	Scale	*M*	*SD*	1	2	3	4	5	6	7	8
1	Age	37.25	10.17	–	.08	-.16	.06	-.29*	-.19	-.21	-.24*
2	Gender	1.49	0.50		–	.10	.00	.10	.11	.07	.11
3	Perceived performance	10.08	2.19			–	-.41***	.41***	.83**	.46**	.63**
4	Nervousness	6.72	2.43				–	-.13	-.46***	-.26*	-.29*
5	Usefullness HMI	8.92	3.49					–	.42***	.30*	.39***
6	Overall trust	9.43	2.36						*.89*	.67**	.77**
7	Initial trust	10.32	2.89							*.87*	.50**
8	Posttask trust	9.22	3.47								*.92*

Reliabilities for multi-item measures are displayed in the diagonal.
Correlations between the variables are based on an n of 72. Male coded as 1; female coded as 2.
**p* < .05; ***p* < .01; ****p* < .001

A.2.3 Overview variables study 3

Table A.16
Scale means, standard deviations, reliabilities, and intercorrelations of scales used in the third user study (N = 18).

	Scale	*M*	*SD*	1	2	3	4	5	6	7	8
1	Age	39.00	9.15	–	.53*	-.07	.01	-.01	-.02	.05	.02
2	Gender	1.50	0.51		–	-.05	-.12	-.06	.00	.13	-.03
3	Perceived performance	11.65	1.83			–	.84***	.94***	.92***	.91***	.96***
4	Trust day 1	10.23	2.06				*.82*	.87***	.80***	.70**	.89***
5	Trust day 2	10.44	2.87					*.97*	.93***	.83***	.97***
6	Trust day 3	10.78	2.76						*.94*	.89***	.97***
7	Trust day 4	11.23	2.71							*.93*	.92***
8	Overall trust	10.79	2.46								*.95*

Reliabilities for multi-item measures are displayed in the diagonal.
Correlations between the variables are based on an n of 18. Male coded as 1; female coded as 2.
$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$