6-18-2022

Driver Trust in Automated Driving Systems

Alexander Stocker
Virtual Vehicle Research Center, alexander.stocker@v2c2.at

DRIVER TRUST IN AUTOMATED DRIVING SYSTEMS

Research Paper

Alexander Stocker, Virtual Vehicle Research GmbH, Graz, Austria, alexander.stocker@v2c2.at

Abstract

Vehicle automation is a prominent example of safety-critical AI-based task automation. Recent digital innovations have led to the introduction of partial vehicle automation, which can already give vehicle drivers a sense of what fully automated driving would feel like. In the context of current imperfect vehicle automation, establishing an appropriate level of driver trust in automated driving systems (ADS) is seen as a key factor for their safe use and long-term acceptance. This paper thoroughly reviews and synthesizes the literature on driver trust in ADS, covering a wide range of academic disciplines. Pulling together knowledge on trustful user interaction with ADS, this paper offers a first classification of the main trust calibrators. Guided by this analysis, the paper identifies a lack of studies on adaptive, contextual trust calibration in contrast to numerous studies that focus on general trust calibration.

Keywords: Trust, automation, trust calibration, automated vehicles, autonomous vehicles.

1 Introduction and Motivation

Enabling fully automated driving is one of the most ambitious automation projects of the present day, and investments in automated driving systems (ADS) have skyrocketed to make transport safer, more efficient, and more comfortable. As a result of the broad media coverage, automated vehicles are on everyone's lips today. While the public perception is that fully automated driving is just around the corner, vehicle manufacturers are pursuing an evolutionary strategy. Increasing the level of vehicle automation should be achieved gradually through advanced driving assistance systems (Fraedrich et al. 2015; Winner et al., 2018; VDA 2020). Recent digital innovations enable partial vehicle automation such as automatically maintaining the lane and keeping the distance. However, using ADS is not about simply activating them. A driver must use current imperfect ADS properly and monitor their performance constantly. Therefore, an *appropriate level of trust* is an important determinant of safe and adequate system use and long-term adoption (cf. Lee and See, 2004; Hoff and Bashir, 2014).

Creating a better understanding of why people adopt or reject technologies has become a widely studied topic, and the importance of trust for technology acceptance research has taken a centre stage in the IS field (Benbasat et al. 2012). Given the rising importance of trust in online environments (Gefen et al., 2008), e-commerce (McKnight et al, 2002; Gefen et al., 2003; Komiak et al., 2006), and e-government (Bélanger and Carter, 2008), IS researchers have shown a significant interest in trust-related works, exploring trust between users and information systems in use (Söllner, 2016). Thereby, IS-related trust research has mainly focused on examining trust relationships between people and interpreted the role of IT artifacts as a means of communication between people (Söllner et al., 2012). However, the increasing automation of artifacts, especially in the context of artificial intelligence (AI), allows them to take over different roles, such as supporting users or taking over complete digital or physical tasks.

For trust to be an important part of a relationship, *individuals must willingly put themselves in vulnerable positions* by delegating responsibility for actions to another party (Lee and See, 2004). As trust is relative to uncertainty, the perception of risk plays a critical role in trust development (Hoff and Bashir, 2014). The use of current imperfect ADS is associated with many risks. ADS can only function within their system boundaries, and perfect automation is still a long way off. For example, the ADS of a production vehicle may automatically switch off in a tight curve, causing the vehicle to leave the ideal lane and leaving the driver only a short time to resume the driving task and return to the ideal lane. Therefore, vehicle manufacturers insist that drivers of ADS-equipped vehicles keep their hands on the steering wheel at all times.

Safety-critical automation systems such as ADS have hardly been the subject of IS research. User trust in safety-critical automation systems is associated with the acceptance of physical vulnerability by their users. Unlike trust in information systems (IS), inappropriate levels of trust in ADS will have more fundamental consequences for the trustor, as over-trust can lead to misuse and probably fatal accidents, while under-trust or mistrust can lead to complete non-use. Hence, *trust in ADS needs to be calibrated appropriately*. If a driver who has great trust in ADS leaves the driving task to an imperfect ADS, that driver may not be alert enough to intervene if the ADS makes a driving error (cf. e.g., Hergeth et al., 2017; Clark and Feng, 2017). Combinations of driver inattention and too excessive trust in the capabilities of imperfect ADS have led to fatal crashes (cf. Banks et al., 2018; Dikmen and Burns, 2017).

Automated driving offers an interesting field of trust-focused research, since AI-based IT artifacts (i.e., the ADS) can, depending on the level of driving automation, relieve the user (i.e., the driver) of complete tasks (i.e., driving a vehicle). However, whenever the vehicle takes over the driving task, it poses a major safety risk to the driver if the ADS does not control the vehicle properly and the driver has too excessive trust in the ADS, as a vehicle accident can result in serious injury. The author would therefore like to highlight the importance of trust as a dynamic concept and the calibration of an appropriate level of trust as a design implication for the IT artifact (i.e., the ADS). While research on driver trust in ADS is still in its infancy, the number of studies on ADS use conducted in vehicle simulators, on test tracks, and on public roads has increased significantly in recent years (cf. e.g., Körber et al., 2018; Payre et al., 2016, Hergeth et al., 2016; Neuhuber et al., 2020). Therefore, the article suggests that IS researchers would benefit from a critical review of studies on driver trust in ADS and addresses the research question of *what influences the calibration of an appropriate level of driver trust in ADS.*

This review paper focuses on *appropriate driver trust* as an important subtopic of trust in automation, illustrates driver trust as a complex and dynamic concept with many interesting implications, and contributes to IS and related literature in several ways. First, the paper provides a critical assessment of an important phenomenon, calibrating an appropriate level of user trust in safety-critical AI-based automation. Addressing this research question can probably kick off a new debate within the IS community on trust in safety-critical AI systems vs. trust in traditional information systems. Second, this review extends the two existing reviews on trust in automation (Lee and See; 2004; Hoff and Bashir, 2014) summarizing the knowledge gained within studies explicitly focused on examining driver trust in ADS. Third, a better knowledge of trust-influencing factors, so-called trust calibrators, will improve human-automation performance and lead to more reliable systems. From a practitioners' view, the derived results offer important implications for the design of ADS.

2 Background

2.1 The rise of automated driving

For decades, ADS such as anti-lock braking systems (ABS) or electronic stability control (ESC) have been supporting drivers in their driving tasks to improve safety and comfort. An increasing and significant market penetration of driver assistance systems has been observed since, which is promoted by Euro NCAP regulations (Euro NCAP 2020). Formerly isolated driver assistance functions are being combined to realize more advanced driver assistance systems with combined longitudinal and lateral guidance (Bengler et al. 2014), which opens up possibilities for a higher degree of vehicle automation (Chan, 2017). The simultaneous use of adaptive cruise control, a system that automatically adjusts vehicle speed based on a driver-defined maximum speed to maintain a safe distance from a vehicle driving ahead, and lane-keeping, a system that keeps the vehicle in the center of the lane, can give drivers a feeling of automated driving, especially when their hands are away from the steering wheel. But drivers do remain responsible for the entire execution of the driving task. Therefore, after a certain period, vehicles with activated ADS usually warn drivers visually (e.g., through text and symbols in the vehicle dashboard), acoustically (e.g., through beeps), and/or haptically (e.g., by steering wheel vibrations) that they must take over vehicle control again when they have taken their hands off the steering wheel. If drivers ignore these warnings, the ADS is deactivated and, on some vehicle models, even a short emergency braking maneuver can be initiated to refocus the attention.

The maturity of ADS can be divided into different levels. Frequently quoted taxonomies for the maturity of automated driving (cf. figure 1) have been released by the Germany Federal Highway Research Institute (BASt), the National Highway Traffic Safety Administration (NHTSA), and the Society of Automotive Engineers (SAE). The widely used SAE J3016 standard (2019) defines *six levels of driving automation*, from no automation (level 0) to full vehicle autonomy (level 5), and serves as the most-cited reference framework. In the first three levels, the human driver has full control, even when ADS are switched on, hands are taken off the steering wheel and feet are taken off the pedals. From level three to level five the human driver has no control over the vehicle when the ADS are engaged. However, at level three the human driver must take over control when the ADS requests the driver to do so, whereas at levels four and five the automatic driving systems will no longer request the human driver to regain control of the vehicle. In levels three and four the ADS can control the vehicle only under limited system conditions and will not operate if these are not met, while in level five the ADS can drive the vehicle under all possible conditions.

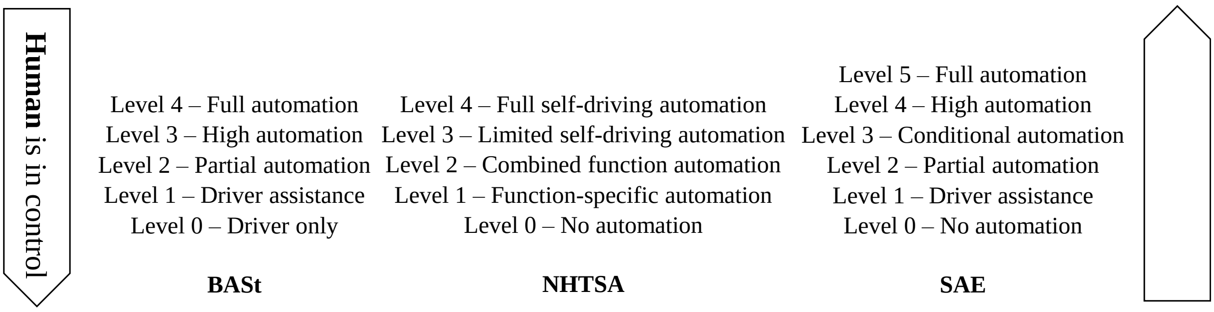

Figure 1. Levels of vehicle automation.

2.2 The concept of trust

The interest in trust began as an investigation of human-human trust. Rousseau et al. (1998) define trust as „a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another". From a social science perspective, Gambetta (2000), defines trust as a "particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before it can monitor such action [..] and in the context in which it affects his own action". If someone is seen as a trusted agent, another agent who trusts that agent is likely to consider collaboration. From an organizational perspective, Mayer et al. (1995) coined "organizational trust" as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other party will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party". From a human factors' perspective, Lee and See (2004) define trust as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability", elaborating this definition to consider context, agent characteristics, and cognitive processes with the 'appropriateness of trust'. Explanations for trust usually contain three components (cf. Hoff and Bashir, 2015). A trustor must give trust to a trustee who must accept the trust, the trustee must have an incentive to perform the task, and there must be a possibility that the trustee will fail to perform the task leading to uncertainty and risk. The trustor is the trusting individual, while the trustee is the individual being trusted. (Mayer et al., 2015). Trust thus describes a relationship that depends on the characteristics of both the trustee and the trustor and the goal-oriented context of the interaction with the automation (Lee and See, 2004).

The formation of trust is a dynamic process because feelings of trust can change if people are exposed to new information (Hoff and Bashir, 2015). Three levels, ability (skills and competencies of the trustee to influence the domain), integrity (the degree to which the trustee adheres to principles which the trustor finds acceptable), and benevolence (the degree to the trustee's intents and motivations are in line with those of the trustor), summarize the base of trust (Mayer et al., 1995; Lee and See, 2004).

2.3 Theoretical lens: Trust in automation

The concept of "trust" has been transferred to the technical domain to better understand interactions between humans and machines, which are becoming increasingly automated. Automation is a technology that selects data, transforms information, makes decisions, and/or controls processes, thereby extending human performance and improving safety (Lee and See, 2004). Automation includes the execution of a function by a machine agent (e.g., a computer, robot, AI, or an ADS) that was previously executed by a human (Parasuraman and Riley, 1997). Just as in interpersonal relationships, trust plays a leading role in determining people's willingness to rely on automation in situations characterized by uncertainty and vulnerability (Hoff and Bashir, 2015). People tend to rely on automation they trust and may reject automation they don't. Therefore, in the context of trust and reliance, trust is an attitude and reliance a behavior (Lee and See, 2004). The complexity of trust can be reduced to three layers of variability (cf. Hoff and Bashir, 2015), dispositional trust (i.e., an individual's overall tendency to trust automation), situational trust (i.e., trust depending on the specific interaction context), and learned trust (trust based on past experiences). In examining the factors that influence trust in automation performance, process and purpose were identified as general bases of trust (Lee and Moray, 1992). An important determinant of humans' choice of manual or automatic system control is their degree of trust (Muir, 1994). Muir (1994) proposed a model for trust in which humans compare their perceptions of persistence, competence, and responsibility with their expectations of automation, and the product of the comparison between perceived performance and expected performance is trust.

The *calibration of an appropriate level of trust* is crucial for avoiding misuse and disuse of the system (Parasuraman and Riley, 1997; Wicks et al., 1999). Trust calibration refers to the correspondence between a person's trust in the automation system and the automation system's capabilities (Lee and Moray, 1994). Since trust is the basis for deciding when to use the ADS and when to drive manually, the level of trust must match the actual capabilities of the ADS (Kraus et al., 2020b). Thereby, Lee and See (2004) distinguish between 'calibrated trust' if trust corresponds to the system capabilities, which leads to appropriate system use, 'overtrust' if trust exceeds the system capabilities, which leads to system abuse, and 'distrust' if trust falls short of the system capabilities, which leads to disuse of the system. In this respect, an appropriate automation use pattern corresponds to a calibrated level of trust, a situation in which the trust level accurately reflects the capabilities of a system and its actual performance (Muir 1994; Kraus et al., 2020b). Misuse and disuse are examples of inappropriate trust that can compromise safety. Appropriate trust and reliance depend on how well the capabilities of the automation are communicated to the user (Lee and See, 2004) and therefore the automation system must be made more trustable. A properly calibrated level of trust is essential for an automated vehicle in which the driver and the ADS must work together as a team (Azevedo-Sa et al., 2020). Drawing parallels from human-human interactions to human-machine interaction, human and machine form a dyad in which trust is a significant factor for system performance. Trust influences the reliance on automation because people react socially to a technology (cf. Lee and See, 2004). Trust is a mediating variable between system properties and the user's assignment decisions (Muir & Moray, 1996). It evolves and adapts over time along with the user's accumulated knowledge of an automated system (Lee and See, 2004).

ADS can be considered as a type of *AI-based automation system*. Recently, the growing discussion on how to facilitate trust in AI systems (cf. Rossi 2018; Meske et al. 2022) and the 'right level of trust' (cf. Banavar 2016; Kalayci et al. 2021) has gained new momentum. Automation systems should integrate explainable AI to make automated decisions more trustworthy to humans. However, despite the rapid progress of ADS, no series vehicle has yet integrated such mechanisms of explainability for drivers and passengers. This is a major research topic currently promoted by the European Commission under the

Partnership for Connected, Collaborative, and Automated Mobility in the Horizon Europe funding program (European Commission 2022).

2.4 Metrics for assessing driver trust in ADS

Trust in automation is very difficult to quantify. As a psychological construct, the only way to assess it directly is through self-reporting with *subjective assessments* (Hergeth et al. (2016b). Therefore, most researchers use subjective assessments before, during, or after experiments, often requiring drivers to perform secondary tasks while using the ADS (e.g., Banks et al., 2018; Lee et al., 2019; Petersen et al., 2019; Walker et al., 2019). For instance, Hartwich et al. (2018) used the unidimensional 12-item trust in automation scale by Jian et al. (2000) to assess trust. Hergeth et al. (2016a) assessed trust by using an adapted 18-item version of the empirically derived scale from Chien et al. (2014) rating items on a 7-point scale. Payre et al. (2016) asked participants randomized questions such as “globally, I trust the automated driving system”, or “I trust the automated driving system to keep distance from a vehicle ahead” in a post-experimental questionnaire using a 7-point Likert-type scale. Körber et al. (2018) proposed a questionnaire with 19 elements divided into the five subscales reliability/competence, familiarity, trust, understanding, and intention of developers, which each containing between two and four items such as “I trust the system”, or “the system might make sporadic errors”.

Trust questionnaires and scales cannot capture temporary changes in trust unless they are collected very frequently (Hergeth et al., 2016b). Therefore, some researchers have investigated whether and how *observable indicators* can be used to infer trust, arguing that the tendency to monitor an automation system is related to trust. More trust in the automation system should lead to less monitoring behavior (Moray, 2000). Gaze behavior and gazing into safety-relevant regions provide a direct measure of automation trust (Hergeth et al. 2016b; Strauch et al. 2019). A combined measurement of gaze and electrodermal activity predicts self-reported trust even better than each of these measurements alone (Walker et al. 2019). Brake and accelerator pedal responses provide a temporal precise indicator for trust (Lee et al., 2019). Exploring real-time estimation of trust by integrating behaviors captured by eye-tracking, system use, and performance on non-driving tasks appears promising (Azevedo-Sa et al. 2020).

3 Methodology

3.1 Review method

To assess the current state of the art the author conducted a qualitative systematic review of the literature (cf. Pare et al., 2015) following common and established guidelines (cf. Webster & Watson, 2002; Okoli and Schabram, 2010; Rowe, 2014). The approach used to identify the relevant literature included a broad sampling frame that covers a wide range of academic disciplines including human factors, psychology, computer science, and information systems that address trust in ADS.

In a first step, the author searched in *Scopus*, a comprehensive, curated abstract and citation database operated by Elsevier and including more than 75 million records and 24.600 titles from 5000 publishers. The author used a keyword-based search with the query “TITLE-ABS-KEY ("trust" AND ("automated driving" OR "autonomous driving" OR "self-driving" OR "automated car" OR "automated vehicle" OR "autonomous car" OR "autonomous vehicle" OR "self-driving car" OR "self-driving vehicle"))” and then limited the search scope to publication title, abstract, and keywords and included publications till November 2020. Since the most important contributions are likely to be found in leading journals (cf. Webster & Watson, 2002), the review process focused exclusively on reviewing journal articles. The reason for limiting the review sample was to ensure quality control of the selected research articles. Journal articles go through a peer-review process that serves as a quality sieve and allows the author to filter out articles that do not meet a certain level of rigor. Searching Scopus returned a preliminary list of 232 articles as input for the systematic literature review. In a second step, the author screened the content of each paper for relevance to the scope of the literature review. Papers were only included in the review sample if they (1) reported the *results of a human-subject experiment*, (2) where *humans directly interacted with an AD*S, and (3) a relationship between *driver trust in the ADS* or a trust-related behavior and at least one another construct was examined (cf. Hoff and Bashir, 2015). Furthermore, the author excluded papers that only refer to the search terms in the abstract but do not make them actual core content. For example, many papers only took advantage of using terms such as automated or autonomous driving to indicate the cutting-edge nature of their presented research (e.g., in the context of trust concerning AI or blockchain in the wider application scenario of automated driving) but did not study driver trust in ADS at all. Work that did not explicitly refer to *driver trust in ADS* but *other trust contexts* were also excluded. For example, journal papers on *public trust in ADS* measured through surveys of citizens (e.g., Liu et al., 2019) were excluded, too. Five papers in the article list were conference papers and were also excluded from the review. The remaining number of relevant journal articles retrieved was 37. In a final step, the author searched the *AIS Electronic Library* and screened the *Senior Scholars' Basket of Journals*, to ensure coverage of the discipline to which the author would like to contribute. However, no journal publications could be identified that met the inclusion criteria. To be more precise, there was no journal publication at all from the IS community on ADS.

3.2 Review sample

The final sample consisted of 37 journal articles, representing a wide range of academic disciplines, but none of the studies were published in IS outlets. For this review, the author provides the journal's Impact Factor (IF), Scimago Journal Rank Indicator (SJR), and the number of peer-reviewed articles. The review includes several human-subject studies from Transportation Research Part F: Traffic Psychology and Behaviour (IF 2.518, SJR Q1, 8 papers), Human Factors (IF 3.165, SJR Q1, 6 papers), Applied Ergonomics (SJR Q1, 4 papers), Transportation Research Part C: Emerging Technologies (IF 6.077, SJR Q1, 3 papers), Cognitive Computations (IF 4.307, SJR Q1, 1 paper), Frontiers in Psychology (IF 2.067, SJR Q1, 1 paper), IEEE Transactions on Human-Machine Systems (IF 3.374, SJR Q1, 1 paper), IEEE Transactions on Intelligent Vehicles (IF 6.319, SJR Q1, 1 paper), International Journal of Human-Computer Studies (IF 3.163, SJR Q1, 1 paper), Journal of Advanced Transportation (IF 1.610, SJR Q1, 1 paper), Journal of Experimental Social Psychology (IF 3.254, SJR Q1, 1 paper), Safety Science (IF 4.105, SJR Q1, 1 paper), Traffic Injury Prevention (IF 1.575, SJR Q1, 1 paper), PlosOne (IF 2.740, SJR Q1, 1 paper), International Journal on Interactive Design and Manufacturing (IF 1.88, SJR Q2, 1 paper), Human Factors and Ergonomics in Manufacturing and Service Industries (IF 1.271, SJR Q2, 1 paper), Multimodal Technologies and Interaction (IF 1.98, SJR Q2, 1 paper), PRESENCE: Virtual and Augmented Reality (IF 0.579, SJR Q3, 1 paper), SAE International Journal of Connected and Autonomous Vehicles (SJR Q3, 1 paper), and Automotive Innovation (1 paper).

Most of the knowledge on driver trust in ADS published stems from driving simulator studies (24), studies on public roads (7), studies on closed test tracks (4), as well as combined driving simulator and test track studies (1) and combined on-road and driving simulator studies (1). The knowledge was gained involving ADS of different automation levels, from level 2 (in simulator studies, test track studies, on-road studies) up to level 5 (in simulator studies and partially in studies on test tracks). Furthermore, the influence of trust or on trust was often examined in connection with other constructs (e.g., Molnar et al., 2018; Ha et al., 2020), practices (e.g., Payre et al., 2016; Banks et al., 2018), and technologies (e.g., Seppelt and Lee, 2019; Ma et al., 2021).

4 Review Results: Driver trust calibrators

In this section, the author presents the results of the qualitative systematic literature review to answer the research question, *what factors influence the calibration of an appropriate level of driver trust in ADS*. The following table 1 lists the main trust calibrators identified. The categorization of trust calibrators is based on the assignment of the single most applicable research topic to a group of related subtopics. For example, subtopics such as different driving styles, different modes of conducting automated maneuvers, adaptive and personalized automation behavior, or decision making in the automatic mode were grouped under the key topic ‘ADS’s driving styles and behavior’. The subtopics assigned to research papers were based on the papers' research foci. Though one research paper can contribute to several subtopics, each research paper was assigned to only one main topic to maintain a simplified and structured classification.

Classes	Trust calibrators	Reviewed papers
Driver/ vehicle user	Personality and usage behavior	Kraus et al. (2020a), Li et al (2020), Banks et al. (2018), Petersen et al. (2019), Molnar et al. (2018), Payre et al. (2016), Hergeth et al. (2016b), Walker et al. (2019)
	Prior knowledge and initial experience	Beggiato and Krems (2013), Körber et al. (2018), Kraus et al. (2020b), Khastgir et al. (2018), Hergeth et al. (2016a),
	General experience of use	Dixit et al (2016), Hartwich et al. (2018), Paddeu et al. (2020), Walker et al. (2018), Wilson et al. (2020), Xu et al. (2018)
Automated driving system (ADS)	ADS feedback	Beller et al. (2013), Koo et al. (2015), Ma et al. (2021), Ha et al. (2020), Seppelt and Lee (2019)
	ADS human-machine interface type	Sonoda and Wada (2017), Wintersberger et al. (2020), Oliveira et al. (2020), Waytz et al. (2014), Niu et al. (2017), Ruijten et al. (2017)
	ADS driving style and behavior	Abe et al. (2017), Banks and Stanton (2016), Kidd et al. (2017), Lee et al. (2019), Ekman et al. (2018), Sun et al. (2020), Strauch et al. (2020)

Table 1. Driver trust calibrators and reviewed author contributions.

4.1 Driver personality and usage behavior

Reviewed literature points to the need to better understand the relationship between driver personality traits and trust calibration. Personality traits are dimensions of individual differences that show a consistent pattern of thoughts, feelings, and behavior. For instance, the five-factor model (FFM), which has become increasingly popular among psychologists, includes the five personality traits, neuroticism, extraversion, openness (to experience), agreeableness, and conscientiousness (cf. Costa and McCrae, 1992). The role of the personality variables expressiveness, self-efficacy, self-esteem, anxiety and locus of control on driver trust was explored by Kraus et al. (2020a), who pointed to the need to better understand the psychological processes by which trust is calibrated before and during the use of ADS. Trust in ADS was significantly predicted by self-esteem, self-efficacy, and state anxiety, underscoring the importance of emotional states and anxiety in building trust when people experience ADS. People with higher openness traits (i.e., people who enjoy variety and novelty, are curious, or seek new experiences) tend to have less trust in ADS, while no significant correlations between trust and the other examined personality traits neuroticism, conscientiousness, agreeableness, and extraversion could be identified (Li et al., 2020). Drivers who stated that they also felt comfortable with other drivers behind the wheel also reported higher trust in ADS (Molnar et al. 2018). Several authors dealt with distinctive trust-relevant driver behavior such as monitoring practices or switching between automated and manual modes. Conducting observations Banks et al. (2018) show that drivers are currently not adequately supported in their monitoring tasks (by the ADS) as they exhibit a behavior that indicates over-trust in the ADS. The higher participants' self-reported trust in ADS, the less they monitored the road, and the more attention they paid to non-driving related secondary tasks (Walker et al. 2019). Situational awareness moderated the effects of trust in ADS, leading to a better performance in secondary tasks (Petersen et al. 2019). In experimental conditions with high situational awareness, subjects waited longer and allowed their automated vehicle to get closer to an approaching stationary vehicle before taking over control, indicating higher levels of trust. There is a consistent relationship between driver trust and gaze behavior, and drivers with higher levels of trust tend to monitor the ADS less frequently (Hergeth et al. 2016b). Unexpected events could cause drivers to react more slowly if they have a higher level of trust (Payre et al. 2016). Hence users of ADS should be trained to improve their control recovery performance in emergencies, and ADS designers should include tutorials with feedback on how to cope in critical situations, as experience improves reaction time performance.

4.2 Prior knowledge about ADS and initial experience

Prior knowledge, especially in connection with first system experience, is an important criterion for trust-building. Trust-promoting and trust-lowering introductory information (i.e., through videos showing the ADS' performance) influence self-reported trust and take-over performance (Körber et al. 2018). Study participants who had higher levels of trust spent less time looking at the road or at the vehicle instruments and more time being engaged in non-driving related tasks. Repeated measurement of trust showed that the experience of takeovers and malfunctions led to a temporary loss of trust, but trust was regained during error-free interaction (Kraus et al. 2020b). A priori information about causes and characteristics of malfunctions eliminates the decrease of trust in case of a malfunction. Knowledge of the ADS capabilities influences driver trust, as the development of trust is a dynamic process and trust needs to be calibrated to the correct levels. The introduction of information about the ADS' true capabilities and limitations increased trust (Khastgir et al. 2018). Informing drivers about ADS safety limits enables them to better calibrate their trust. When examining the effects of prior familiarization with ADS take-over requests, Hergeth et al. (2016a) found that participants who were not familiar reported the highest trust before and after experiencing the ADS. On average the involved respondents had rather a low trust in ADS and were more concerned about giving up control to the ADS. Hence, prior familiarization with take-over requests from the ADS (i.e., the importance of familiarity with critical limitations of an automation system) facilitates trust calibration, particularly during initial ADS use. Exploring the effect of divergent initial mental models of ADS on trust, Beggiato and Krems (2013) show that ADS errors do not negatively affect trust if they are known beforehand.

4.3 ADS general experience of use

The amount of general user experience with ADS increases driver trust (Xu et al., 2018). Driver trust is a positive predictor of the driver's intention to use ADS and the drivers' willingness to reuse ADS. Hartwich et al. (2018) conducted studies with two different age groups, showing that both age groups a priori showed slightly positive trust, which increased significantly after initial ADS experiences and remained stable afterward. Paddeu et al. (2020) examined how trust is influenced by certain attributes of the driving experience, such as speed and gaze direction, and found that individuals in an autonomous shared vehicle had more trust when facing forward and driving at slower speeds, while both trust ratings increased after the driving experience. While researchers confirmed the important relationship between driver trust and general driving experience in laboratory settings, the literature also aims for confirmation in on-road studies. For instance, Wilson et al. (2020) carried out an on-road study of level 2 vehicles showing a statistically significant increase in the trust after the drives. The main reasons for trust evolvement as mentioned by drivers were the evidence of the ADS reacting to other traffic and environment as well as the dynamic visual display that showed nearby vehicles and road infrastructure. However, participants observed situations where they mistakenly thought they were in automated mode, a phenomenon known as mode confusion. Examining the impact of experiencing level 2 vehicles on the road, too, Walker et al. (2018) showed significant changes in self-reported trust following the driving experiences: While drivers first overestimated ADS capabilities before experiencing them on the road, they had a better understanding afterward, which led to better trust calibration. Analyzing data released from the California automated vehicle trials, Dixit et al (2016) showed that the number of autonomous miles traveled correlates with the number of accidents observed, while reaction times increased with the number of kilometers traveled, which may indicate that trust in ADS increases with increasing mileage.

4.4 ADS feedback to the driver

Direct feedback about system boundaries and the proper explanation of ADS maneuvers to the driver has a direct impact on trust. For instance, Koo et al. (2015) investigated the impact of driver feedback in a simulator study in which different types of messages from the automation system to the driver were tested, showing that messages describing the justification for automated actions (such as "obstacle ahead") are preferred by drivers, lead to better driving performance, and create a higher level of trust. Investigating whether visual feedback can influence driver trust, and what level of feedback produces the appropriate level of trust, Ma et al. (2021) showed that the participant group experiencing high visual feedback gave the highest trust ratings. However, once driver trust is calibrated to an appropriate level, participants may even desire a reduction of the level of visual feedback. Probing the costs and benefits of providing continuous feedback to drivers on the limits and system behavior of imperfect ADS, Seppelt and Lee (2019) pointed out that vehicle dashboards should better inform about the situation-specific behavior of the ADS, rather than simply alerting drivers to errors and the need to resume vehicle control. However, the latter is the current state of the art in serial vehicles. With a more accurate mental model, i.e., a better understanding by the driver of what the ADS contains and why and how it works the way it does (cf. Carroll and Olson, 1988), drivers were able to build appropriate confidence based on a clearer understanding of the automation behavior at the moment when the ADS reached operational limits. Examining the effects of different explanation types of the status of ADS and perceived risk on trust in ADS, Ha et al. (2020) showed that attributional explanations affect trust most: At low perceived risk, attributional explanations describing why and how the ADS acted (e.g., stopping the vehicle after identifying the sudden appearance of a pedestrian on the road) were most effective in increasing trust, while at high perceived risk attributional explanations worsened the situation. Presenting information about automation uncertainty increased trust in the event of an automation error, and the ADS that displayed an uncertainty symbol to the driver received the highest trust score (Beller et al. 2013).

4.5 ADS human-machine interface type

Several authors pointed out the important role of the ADS' human-machine interface. For instance, Sonoda and Wada (2017) examined driver trust using vibrotactile displays such as wristbands with a motor attached: The display of spatial information such as close traffic objects by a haptic stimulus is effective in increasing driver trust. Modern ADS can take advantage of head-up displays and augmented reality (AR) technologies. Wintersberger et al. (2020) explored the potential of AR to increase driver trust through communicating ADS decisions in the driver's field of vision. Results suggest that the augmentation of traffic objects and participants otherwise invisible (e.g., because of fog or snowfall) or the display of upcoming driving maneuvers is feasible to increase trust. Oliveira et al. (2020) showed that the use of AR windshields entailed the highest level of trust, but drivers wonder if they need to see information all the time for appropriate trust calibration. Anthropomorphism, i.e., the degree to which humans ascribe human-like characteristics to non-human agents has been studied in the context of driver trust, too. For instance, Waytz et al. (2014) examined the extent to which a non-human agent is anthropomorphized with a human-like mind in trust in ADS research, showing that participants trusted the ADS to perform more competently as it acquired more anthropomorphic features. Investigating the effects of the anthropomorphic embodiment of information about the ADS` driving maneuvers on trust (e.g., when the automated vehicle is going to turn left, visualized eyes were looking to the left), Niu et al. (2017) found that a combination of symbolic and anthropomorphic information led to significantly higher trust than symbolic information alone. Ruijten et al (2017) gave vehicles a human voice and simulated intelligent conversations, showing that interfaces better mimicking human behavior in explaining decisions (e.g., "we are on a cobblestone road with pedestrians, I am slowing down") may help increase driver trust.

4.6 ADS driving styles and automation behavior

Several authors identified a relationship between driver trust and the ADS' driving style and behavior: For example, Abe et al. (2017) simulated automatic overtaking of scooters and bicycles in a study and showed a direct influence of speed, lateral distance to the object, and start timing of the automatic steering maneuver on driver trust. Banks and Stanton (2016) examined the idea of driver-initiated automation, where the ADS only provides decision support than can be accepted or ignored, showing that drivers need a clear understanding of what the ADS can do to promote driver trust. Unexpected lane changes and unsafe lane offers negatively impact driver trust. Kidd et al. (2017) probed driver trust with different level 2 vehicles in a multi-week on-road study showing that trust varied between different ADS and among different types of vehicles, indicating the lowest trust for the lane-keeping system, a central function of ADS. Furthermore, authors have investigated the influence of concrete automated driving styles. For instance, Lee et al. (2019) exposed drivers to a fully automated vehicle with three different driving styles, aggressive, moderate, and conservative, asking drivers to indicate their dissatisfaction with the ADS by pressing the brake or the accelerator pedal, indicating pedal inputs as a temporally precise measure of driver trust. A defensive automated driving style was perceived as more trustworthy, also because it is more predictable by users (Ekman et al., 2018). ADS users have, on average, higher trust in a personalized automation mode than in the standard automation mode or the manual driving mode, as personalization makes it easier for participants to judge the quality of the ADS (Sun et al., 2020). Categorizing passengers' eye movements as safety-relevant or not safety-relevant, Strauch et al. (2020) suggest that the ADS' driving style affects trust: being driven fully autonomously led to a lower self-reported trust than believing to be driven by a human driver.

5 Discussion and conclusion

5.1 Theoretical contributions

The paper draws on a critical literature review, assessing the current state of research on driver trust in ADS. Using trust in automation as a theoretical lens, the paper identifies six trust calibrators for adaptive, context-dependent trust calibration (cf. Figure 2) each relating to two classes, the human (driver) and the ADS (the safety-critical system), determining the evolution of trust during (imperfect) ADS use.

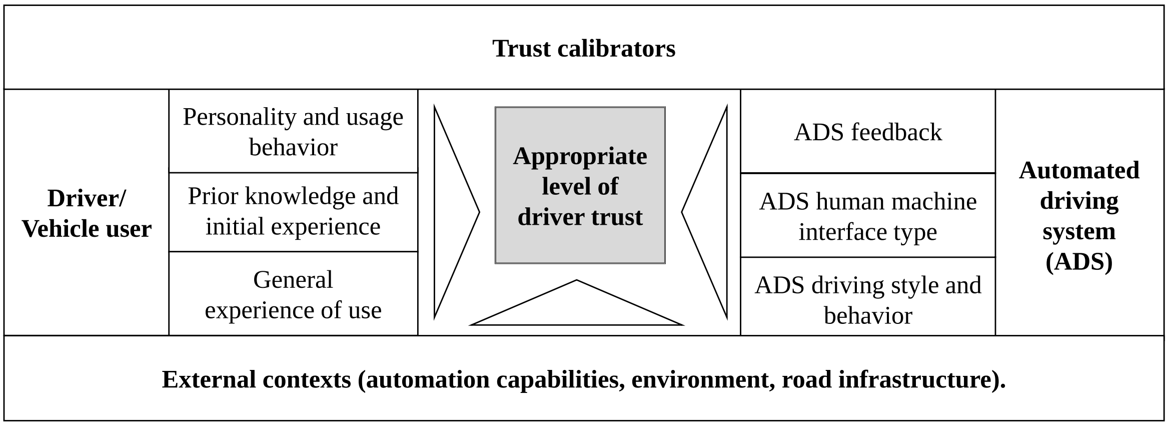

Figure 2. Adaptive, context-dependent trust calibration

The key contribution of this paper lies in *clustering and synthesizing recent progress in trust in ADS* for IS researchers, identifying the main trust calibrators from the literature. In each of the studies, one or more factors relevant to driver trust were examined, and experimental conditions were created that could increase trust. To give an example, drivers may be slower to respond when taking over the driving task, if they have a generally high level of trust in ADS (Payre et al 2016) as they might focus their attention on non-driving related tasks. However, drivers with certain personality traits, such as higher openness, may tend to trust the ADS less and therefore use the ADS less intensively (Li et al 2020). An inappropriate level of trust, i.e., both too high and too low, can therefore have negative consequences for drivers. The paper builds upon two literature reviews by Lee and See (2004) and Hoff and Bashir (2015) that focus on trust in automation in general, but not on trust in ADS. It summarizes and synthesizes research published in academic journals from human-subjects experiments in which humans directly interacted with ADS. Thereby, the paper *extends both reviews by integrating the most recent research on trust in automation* from 2015 to 2021 that could not be included in those two reviews due to the date of publication and exclusively focuses on one type of trust-relevant automation, ADS. Therefore, this work is the first systematic literature review that explicitly focuses on driver trust in automated driving.

Second, the paper highlights, why creating a *better understanding of what contributes to calibrating trust* is important, given the role that trust plays in the use and misuse of ADS as ADS may be over-relied upon, used in unintended ways, or not used at all. For instance, the design of the ADS' human-machine interface and feedback mechanism on system boundaries and decisions influences the level of driver trust (e.g., Seppelt and Lee, 2019; Sonoda and Wada, 2017; Koo et al., 2015) as they may affect the driver's understanding of the ADS' capabilities. When the ADS encounters a situation that it cannot properly handle, it can provide appropriate feedback to the driver in advance and properly calibrate (i.e., also decrease) driver trust so that the driver takes control of the vehicle on time - rather than forcing a risky situation by over-relying on the imperfect ADS. However, most of the literature reviewed focuses on the core mechanisms for increasing driver trust, which is generally considered beneficial. But with imperfect automation, too much trust in automation can have negative impacts.

Third, the paper aims to highlight the *difference between research on trust in ADS in general and research on the calibration of an appropriate level of trust* depending on the user and system context. While people should generally trust ADS sufficiently to use them in the first place (i.e., trust is an important determinant of user acceptance), people's trust in automation should be understood as a dynamic concept that may stimulate correct system use and prevent inadequate and unsafe system use. The proper calibration of driver trust in the ability of the ADS to handle a given situation depends not only on the external context (automation capabilities, environment, road infrastructure) but also on the presented trust calibrators and is crucial for road safety.

Fourth, most research reviewed that addresses trust calibration contributes to the first paradigm, general trust calibration, while *only a few studies address adaptive, context-dependent trust calibration*. To calibrate an appropriate level of driver trust, the driver must be able to correctly assess the abilities, behavior, and limitations of the ADS at all times. The paper addresses the problem of inadequate trust by arguing that trust in ADS should be calibrated by closely aligning trust with the ADS' ability to handle specific driving situations. Herein also lies an important distinction between trust in ADS and acceptance of ADS. A driver may, in general, accept ADS as a useful technology and use it. However, a driver may have a sufficient level of trust in ADS to correctly handle a particular driving situation (e.g., when driving on a motorway with clearly visible center and boundary lines in sunny weather) but may distrust the ADS to handle another one (e.g., driving on a winding country road in the rain with poorly visible side boundary lines). Therefore, the level of trust must always be calibrated to a particular automation context. Supporting over-trust should not be a design goal for trustworthy (vehicle) automation systems. Almost all studies reviewed address either driver trust or general trust calibration, and there is a lack of studies on adaptive, contextual trust calibration. However, contextual factors such as the ADS' capabilities, the environment (e.g., light and weather conditions) and the road infrastructure can have a significant impact on trust calibration, too.

5.2 Practical implications

The six identified trust calibrators can help close the current gap in realizing higher levels of vehicle automation. For example, prior knowledge about ADS is critical for trust calibration. Hence communicating proper knowledge to future users (e.g., through vehicle manuals or special training) and teaching them how to use ADS appropriately are important long-term strategies. Furthermore, the research offers suggestions for ADS designers in terms of user interface technologies, driver feedback approaches, or the characteristics of the driving styles coded into the automation. Finally, the research findings may suggest ways to make ADS more trustworthy and accepted by drivers.

5.3 Limitations and future research

The limitations of this study may serve as a first starting point for future research. Since driver trust in ADS is a relatively new area of research, some studies reviewed may lack theoretical foundations. Although the author used Scopus as the main source and starting point for the literature review, a search within the AIS Electronic Library and the Senior Scholars' Basket of Journals was conducted, too. The majority of peer-reviewed papers in the review sample were published in academic journals with high reputations in human factors and transport-related research, all of which are located in Q1 in the SCR journal ranking, which includes the top 25% journals indexed. Two peer-reviewed papers were in Q2, one in Q3, and one was not included in the SJR ranking due to the novelty of the journal. Some of the academic journals considered, although very prestigious, may still be unknown to IS researchers as they come from other fields, which may reduce IS researchers’ trust in their contributions.

Through searching the AIS library, the paper reveals that *automated driving has not yet generated much interest in the IS community*, which raises the question of whether the IS community may be overlooking an important digital innovation in the context of AI-based safety-critical automation systems. There have been some publications on the acceptance and adoption of automated driving ins IS conference proceedings (cf. e.g., Hein et al. 2018; Lackes et al. 2018), but the evolution of trust (during ADS use) is not studied. However, the author would like to emphasize that ADS offers an interesting field for such trust-focused research relevant to the IS community, too. Unlike automation research, previous research about trust in IT/IS did not address inappropriate trust and (adaptive, contextual) trust calibration, while focusing on general-purpose IT and assuming that the higher the trust, the higher the usage (Chen et al. 2021). Unlike trust in IT/IS, an inappropriate level of trust in ADS can have more fundamental consequences for the trustor due to the higher vulnerability. Over-trust in ADS can lead to misuse and even fatal accidents (cf. e.g., Banks et al., 2018; Dikmen and Burns, 2017), while under-trust can lead to the complete non-use of these important safety features. ADS can provide an interesting research context for the IS community in which theory development and large-scale empirical studies in the use of AI-enabled safety-critical systems can be conducted. A key factor that distinguishes automated driving from other digital technologies being currently studied by the IS community is the *prioritization of calibrating user trust to an appropriate level*, as neither too much nor too little trust leads to safe system use. Therefore, the author calls on the IS community to adopt the context of automated driving as a proxy for AI-based safety-critical automation systems and gain new theoretical insights.

5.4 Concluding remarks

The idea of automated vehicles is by no means new. However, enabling fully automated (autonomous) driving is one of the most challenging automation projects, if not the most challenging at all. Research into factors that influence the calibration of an appropriate level of trust in ADS is essential for safe use and acceptance of imperfect ADS. This review extracts six main trust calibrators from the literature and synthesizes recent progress in driver trust research. The results suggest two classes of trust calibrators. The first class relates to the ADS user and includes the personality and behavior of the user, initial experience with the ADS, and general long-term experience with the ADS, while the second class relates to the automation system and includes the ADS’ feedback to the driver, the ADS’ type of human-machine interface and the ADS’ driving style and behavior. Almost all reviewed studies address either driver trust or general trust calibration while there is a *lack of studies on adaptive, contextual trust calibration*. The paper, therefore, calls for more empirical studies that focus on adaptive, context-dependent trust calibration in the use of automation systems.

Acknowledgment

This research was funded by the program “Mobilität der Zukunft” of the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation, and Technology (BMK) under project number 884333 (FFG project UT4AD).

References

Abe, G., Sato, K. and Itoh, M. (2017). Driver trust in automated driving systems: The case of overtaking and passing. *IEEE Transactions on Human-Machine Systems*, 48(1), pp.85-94.

Azevedo-Sa, H., Jayaraman, S.K., Esterwood, C.T., Yang, X.J., Robert, L.P. and Tilbury, D.M. (2020). Real-time estimation of drivers' trust in automated driving systems. *International Journal of Social Robotics*, pp.1-17.

Banavar, G. (2016). Learning to trust artificial intelligence systems. *Report, IBM, Armonk, NY*.

Banks, V.A., Eriksson, A., O'Donoghue, J. and Stanton, N.A. (2018). Is partially automated driving a bad idea? Observations from an on-road study. *Applied ergonomics*, *68*, pp.138-145.

Banks, V.A., Plant, K.L. and Stanton, N.A. (2018). Driver error or designer error: Using the Perceptual Cycle Model to explore the circumstances surrounding the fatal Tesla crash on 7th May 2016. *Safety science*, *108*, pp.278-285.

Beggiato, M. and 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*, pp.47-57.

Bélanger, F. and Carter, L. (2008). Trust and risk in e-government adoption. *The journal of strategic information systems*, *17*(2), pp.165-176.

Beller, J., Heesen, M. and Vollrath, M. (2013). Improving the driver–automation interaction: An approach using automation uncertainty. *Human factors*, *55*(6), pp.1130-1141.

Bengler, K., Dietmayer, K., Farber, B., Maurer, M., Stiller, C. and Winner, H. (2014). Three decades of driver assistance systems: Review and future perspectives. *IEEE Intelligent transportation systems magazine*, *6*(4), pp.6-22.

Carroll, J.M. and Olson, J.R. (1988). Mental models in human-computer interaction. *Handbook of human-computer interaction*, pp.45-65.

Chan, C.Y. (2017). Advancements, prospects, and impacts of automated driving systems. *International journal of transportation science and technology*, *6*(3), pp.208-216.

Chen, Y., Zahedi, F.M., Abbasi, A. and Dobolyi, D. (2021). Trust calibration of automated security IT artifacts: A multi-domain study of phishing-website detection tools. *Information & Management*, *58*(1), p.103394.

Clark, H. and Feng, J. (2017). Age differences in the takeover of vehicle control and engagement in non-driving-related activities in simulated driving with conditional automation. *Accident Analysis & Prevention*, *106*, pp.468-479.

Costa Jr, P.T. and McCrae, R.R. (1992). Revised NEO personality inventory (NEO-PI-R) and NEO five-factor (NEO-FFI) inventory professional manual. *Odessa, Fl: PAR*.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pp.319-340.

Dikmen, M. and Burns, C. (2017). Trust in autonomous vehicles: The case of Tesla Autopilot and Summon. In *2017 IEEE International conference on systems, man, and cybernetics (SMC)* (pp. 1093-1098). IEEE.

Dixit, V.V., Chand, S. and Nair, D.J. (2016). Autonomous vehicles: disengagements, accidents and reaction times. *PLoS one*, *11*(12), p.e0168054.

Ekman, F., Johansson, M., Bligård, L.O., Karlsson, M. and Strömberg, H. (2019). Exploring automated vehicle driving styles as a source of trust information. *Transportation research part F: traffic psychology and behaviour*, *65*, pp.268-279.

Euro NCAP (2020). "How To Read The Stars", https://www.euroncap.com/en/about-euro-ncap, accessed on 15. April 2021.

European Commission (2020). EU Road Safety Policy Framework 2021-2030, Next step towards 'Vision Zero', op.europa.eu/en/publication-detail/-/publication/d7ee4b58-4bc5-11ea-8aa5-01aa75ed71a1, accessed on 15. April 2021.

European Commission (2022). Connected, Cooperative and Automated Mobility - CCAM Partnership – www.ccam.eu accessed 11 03 2022.

Fraedrich, E., Beiker, S. and Lenz, B. (2015). Transition pathways to fully automated driving and its implications for the sociotechnical system of automobility. *European Journal of Futures Research, 3*(1), pp.1-11.

Gambetta, D. (2000). Can we trust trust. *Trust: Making and breaking cooperative relations, 13*, pp.213-237.

Gefen, D., Benbasat, I. and Pavlou, P. (2008). A research agenda for trust in online environments. *Journal of Management Information Systems*, *24*(4), pp.275-286.

Gefen, D., Karahanna, E. and Straub, D.W. (2003). Trust and TAM in online shopping: An integrated model. *MIS quarterly*, pp.51-90.

Ha, T., Kim, S., Seo, D. and Lee, S. (2020). Effects of explanation types and perceived risk on trust in autonomous vehicles. *Transportation research part F: traffic psychology and behaviour*, *73*, pp.271-280.

Hartwich, F., Witzlack, C., Beggiato, M. and Krems, J.F. (2019). The first impression counts–A combined driving simulator and test track study on the development of trust and acceptance of highly automated driving. *Transportation research part F: traffic psychology and behaviour*, *65*, pp.522-535.

Hein, D., Rauschnabel, P., He, J., Richter, L. and Ivens, B. (2018). What drives the adoption of autonomous cars? Thirty ninth International Conference on Information Systems, San Francisco.

Hergeth, S., Lorenz, L. and Krems, J.F. (2017). Prior familiarization with takeover requests affects drivers' takeover performance and automation trust. *Human factors*, *59*(3), pp.457-470.

Hergeth, S., Lorenz, L., Vilimek, R. and Krems, J.F. (2016). Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving. *Human factors*, *58*(3), pp.509-519.

Hoff, K.A. and Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, *57*(3), pp.407-434.

Jian, J.Y., Bisantz, A.M. and Drury, C.G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International journal of cognitive ergonomics*, *4*(1), pp.53-71.

Kalayci, T.E., Kalayci, E.G., Lechner, G., Neuhuber, N., Spitzer, M., Westermeier, E. and Stocker, A. (2021). Triangulated investigation of trust in automated driving: Challenges and solution approaches for data integration. *Journal of Industrial Information Integration*, *21*, p.100186.

Kidd, D.G., Cicchino, J.B., Reagan, I.J. and Kerfoot, L.B. (2017). Driver trust in five driver assistance technologies following real-world use in four production vehicles. *Traffic injury prevention*, *18*(sup1), pp.S44-S50.

Khastgir, S., Birrell, S., Dhadyalla, G. and Jennings, P. (2018). Calibrating trust through knowledge: Introducing the concept of informed safety for automation in vehicles. *Transportation research part C: emerging technologies*, *96*, pp.290-303.

Komiak, S.Y. and Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS quarterly*, pp.941-960.

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L. and Nass, C. (2015). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *9*(4), pp.269-275.

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

Kraus, J., Scholz, D., Messner, E.M., Messner, M. and Baumann, M. (2020). Scared to trust?–predicting trust in highly automated driving by depressiveness, negative self-evaluations and state anxiety. *Frontiers in Psychology, 10*, p.2917.

Kraus, J., Scholz, D., Stiegemeier, D. and Baumann, M. (2020). The more you know: trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction, and system transparency. *Human factors*, *62*(5), pp.718-736.

Lackes, R., Siepermann, M. and Vetter, G. (2020). Where can I take you?–The drivers of autonomous driving adoption. *European Conference on Information Systems*. 159.

Lee, J.D., Liu, S.Y., Domeyer, J. and DinparastDjadid, A. (2021). Assessing drivers' trust of automated vehicle driving styles with a two-part mixed model of intervention tendency and magnitude. *Human factors*, *63*(2), pp.197-209.

Lee, J.D. and See, K.A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, *46*(1), pp.50-80.

Li, W., Yao, N., Shi, Y., Nie, W., Zhang, Y., Li, X., Liang, J., Chen, F. and Gao, Z. (2020). Personality Openness Predicts Driver Trust in Automated Driving. *Automotive Innovation*, *3*(1), pp.3-13.

Liu, P., Yang, R. and Xu, Z. (2019). Public acceptance of fully automated driving: Effects of social trust and risk/benefit perceptions. *Risk Analysis*, *39*(2), pp.326-341.

Ma, R.H., Morris, A., Herriotts, P. and Birrell, S. (2021). Investigating what level of visual information inspires trust in a user of a highly automated vehicle. *Applied Ergonomics*, *90*, p.103272.

Mayer, R.C., Davis, J.H. and Schoorman, F.D. (1995). An integrative model of organizational trust. *Academy of management review*, *20*(3), pp.709-734.

McKnight, D.H., Choudhury, V. and Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), pp.334-359.

Meske, C., Bunde, E., Schneider, J. and Gersch, M. (2022). Explainable artificial intelligence: objectives, stakeholders, and future research opportunities. *Information Systems Management*, *39*(1), pp.53-63.

Molnar, L.J., Ryan, L.H., Pradhan, A.K., Eby, D.W., Louis, R.M.S. and Zakrajsek, J.S. (2018). Understanding trust and acceptance of automated vehicles: An exploratory simulator study of transfer of control between automated and manual driving. *Transportation research part F: traffic psychology and behaviour*, *58*, pp.319-328.

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), pp.1905-1922.

Neuhuber, N., Lechner, G., Kalayci, T.E., Stocker, A. and Kubicek, B. (2020). Age-related differences in the interaction with advanced driver assistance systems-a field study. In *International Conference on Human-Computer Interaction* (pp. 363-378). Springer, Cham.

Niu, D., Terken, J. and Eggen, B. (2018). Anthropomorphizing information to enhance trust in autonomous vehicles. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *28*(6), pp.352-359.

Okoli, C. and Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research, Sprouts: Working Papers on Information Systems, 10 (26). http://sprouts.aisnet.org/10-26.

Oliveira, L., Burns, C., Luton, J., Iyer, S. and Birrell, S. (2020). The influence of system transparency on trust: Evaluating interfaces in a highly automated vehicle. *Transportation research part F: traffic psychology and behaviour*, *72*, pp.280-296.

Paddeu, D., Parkhurst, G. and Shergold, I. (2020). Passenger comfort and trust on first-time use of a shared autonomous shuttle vehicle. *Transportation Research Part C: Emerging Technologies*, *115*, p.102604.

Parasuraman, R. and Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors*, *39*(2), pp.230-253.

Paré, G., Trudel, M.C., Jaana, M. and Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, *52*(2), pp.183-199.

Payre, W., Cestac, J. and Delhomme, P. (2016). Fully automated driving: Impact of trust and practice on manual control recovery. *Human factors*, *58*(2), pp.229-241.

Petersen, L., Robert, L., Yang, J. and Tilbury, D. (2019). Situational awareness, driver's trust in automated driving systems and secondary task performance. *SAE International Journal of Connected and Autonomous Vehicles, Forthcoming*.

Rossi, F. (2018). Building trust in artificial intelligence. *Journal of international affairs*, *72*(1), pp.127-134.

Rowe, F. (2014). What literature review is not: diversity, boundaries and recommendations. *European Journal of Information Systems*, *23*(3), pp.241-255.

Rousseau, D.M., Sitkin, S.B., Burt, R.S. and Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of management review*, *23*(3), pp.393-404.

Ruijten, P.A., Terken, J. and Chandramouli, S.N. (2018). Enhancing trust in autonomous vehicles through intelligent user interfaces that mimic human behavior. *Multimodal Technologies and Interaction*, *2*(4), p.62.

SAE (2019). SAE Standards News: J3016 automated-driving graphic update, www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic, accessed 13. April 2021.

Seppelt, B.D. and Lee, J.D. (2019). Keeping the driver in the loop: Dynamic feedback to support appropriate use of imperfect vehicle control automation. *International Journal of Human-Computer Studies*, *125*, pp.66-80.

Seppelt, B.D. and Victor, T.W. (2016). Potential solutions to human factors challenges in road vehicle automation. In *Road vehicle automation 3* (pp. 131-148). Springer, Cham.

Söllner, M., Hoffmann, A. and Leimeister, J.M. (2016). Why different trust relationships matter for information systems users. *European Journal of Information Systems*, *25*(3), pp.274-287.

Söllner M., Hoffmann A., Hoffmann H., Wacker A., Leimeister J.M. (2012). „Understanding the formation of trust in IT artifacts", *International Conference on Information Systems*.

Sonoda, K. and Wada, T. (2017). Displaying system situation awareness increases driver trust in automated driving. *IEEE Transactions on Intelligent Vehicles*, *2*(3), pp.185-193.

Strauch, C., Mühl, K., Patro, K., Grabmaier, C., Reithinger, S., Baumann, M. and Huckauf, A. (2019). Real autonomous driving from a passenger's perspective: Two experimental investigations using gaze behaviour and trust ratings in field and simulator. *Transportation research part F: traffic psychology and behaviour*, *66*, pp.15-28.

Sun, X., Li, J., Tang, P., Zhou, S., Peng, X., Li, H.N. and Wang, Q. (2020). Exploring personalised autonomous vehicles to influence user trust. *Cognitive Computation*, *12*(6), pp.1170-1186.

Venkatesh, V., Morris, M.G., Davis, G.B. and Davis, F.D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, pp.425-478.

Walker, F., Boelhouwer, A., Alkim, T., Verwey, W.B. and Martens, M.H., (2018). Changes in trust after driving level 2 automated cars. *Journal of advanced transportation*, *2018*.

Walker, F., Wang, J., Martens, M.H. and Verwey, W.B. (2019). Gaze behaviour and electrodermal activity: Objective measures of drivers' trust in automated vehicles. *Transportation research part F: traffic psychology and behaviour*, *64*, pp.401-412.

Waytz, A., Heafner, J. and Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, pp.113-117.

Webster, J. and Watson, R.T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, pp.xiii-xxiii.

Wicks, A.C., Berman, S.L. and Jones, T.M. (1999). The structure of optimal trust: Moral and strategic implications. *Academy of Management review*, *24*(1), pp.99-116.

Wilson, K.M., Yang, S., Roady, T., Kuo, J. and Lenné, M.G. (2020). Driver trust & mode confusion in an on-road study of level-2 automated vehicle technology. *Safety Science*, *130*, p.104845.

Winner, H., Wachenfeld, W. and Junietz, P. (2018). Validation and introduction of automated driving. In *Automotive Systems Engineering II* (pp. 177-196). Springer, Cham.

Wintersberger, P., Frison, A.K., Riener, A. and Sawitzky, T.V. (2019). Fostering user acceptance and trust in fully automated vehicles: Evaluating the potential of augmented reality. *PRESENCE: Virtual and Augmented Reality*, *27*(1), pp.46-62.

Xu, Z., Zhang, K., Min, H., Wang, Z., Zhao, X. and Liu, P. (2018). What drives people to accept automated vehicles? Findings from a field experiment. *Transportation research part C: emerging technologies*, *95*, pp.320-334.