How the initial level of trust in automated driving impacts drivers' behaviour and early trust construction

J.-B. Manchon, Mercedes Bueno, Jordan Navarro

How the initial level of Trust in Automated Driving impacts drivers' behaviour and early trust construction

J. B. Manchon[1,2], **Mercedes Bueno**[1], and **Jordan Navarro**[2]

[1] VEDECOM Institute, Versailles, France

[2] Laboratoire d'Etude des Mécanismes Cognitifs (EA 3082), University Lyon 2, Bron, France

Correspondence address: J. B. Manchon, Institut VEDECOM, 23 bis Allée des Marronniers, 78000 Versailles, France; e-mail: jb.manchon@vedecom.fr.

Word count: 6,242

ABSTRACT

Trust in Automation is known to influence human-automation interaction and user behaviour. In the Automated Driving (AD) context, studies showed the impact of drivers' Trust in Automated Driving (TiAD), and linked it with, e.g., difference in environment monitoring or driver's behaviour. This study investigated the influence of driver's initial level of TiAD on driver's behaviour and early trust construction during Highly Automated Driving (HAD). Forty drivers participated in a driving simulator study. Based on a trust questionnaire, participants were divided in two groups according to their initial level of TiAD: high (Trustful) vs. low (Distrustful). Declared level of trust, gaze behaviour and Non-Driving-Related Activities (NDRA) engagement were compared between the two groups over time. Results showed that Trustful drivers engaged more in NDRA and spent less time monitoring the road compared to Distrustful drivers. However, an increase in trust was observed in both groups. These results suggest that initial level of TiAD impact drivers' behaviour and further trust evolution.

Keywords: trust in automation, automated driving, driver's behaviour

1. Introduction

Automated driving is becoming part of traffic flows, and such technology raises new questions. As automation capabilities progress, driver-vehicle interaction and cooperation evolve (Navarro, 2019). For example, highly automated driving (HAD, level 4, SAE, 2016) provides full control of the driving task, allowing drivers to focus entirely on non-driving-related activities (NDRA), since the vehicle can perform a minimum-risk manoeuvre if the driver is unable to resume manual control.

Trust is a key element that influences human-machine interaction (HMI) in many ways (Lee & See, 2004; Parasuraman & Riley, 1997), deeply impacting human-system overall performance (Lee & Moray, 1992), and being a determinant of automation usage as well (Parasuraman & Riley, 1997; Schaefer et al., 2016). Trust in automation (TiA) is commonly defined as "*the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability*" (Lee & See, 2004). TiA has been widely studied during recent decades among automation experts (e.g., plane pilots, power plant supervisors; see Parasuraman & Riley, 1997) as a major reliance factor (e.g., Lee & See, 2004; Sheridan, 2019). Drivers have very diverse profiles in terms of level of experience, abilities, and mental and physical condition. They also have various expectations related to HAD that are a product of biased mental models created by cultural elements such as advertising, misconceptions, and hearsay. Furthermore, driving contexts are also varied and highly changing. This combination of factors seems to support a relatively new paradigm which leads to considering trust in automated driving (TiAD) as a different, although closely related, construct from TiA (Manchon et al., 2020).

TiAD is defined here as the attitude that a driver has about HAD, which allows drivers to delegate the driving task to automation to improve safety and comfort, although the risk of accidents continues to exist.

TiA seems to be high when operators' self-confidence is low, but low in the opposite situation (Lee & Moray, 1994). Most drivers might be prone to misjudging their driving skills, leading to poorly calibrated TiAD (Wintersberger & Riener, 2016), and this raises concerns about trust calibration during their interactions with the HAD system. Trust calibration describes the conformity of an operator's TiA with actual automation capabilities (Lee & See, 2004). When this calibration is not correct, two types of situations may develop. For example, excessive trust could lead to hazardous conditions when using this kind of system in traffic (Hancock et al., 2011; Hoff & Bashir, 2015; Parasuraman & Riley, 1997), as has been shown by the many accidents due to poor advanced driver-assistance systems (ADAS) monitoring (e.g., NHTSA, 2017). On the other hand, distrust might be equally dangerous in specific circumstances, such as in the case of a driver who maintained manual control while experiencing drowsiness. It is, therefore, important to understand drivers' trust-calibration process in order to design well-accepted and safer automated driving systems that inspire a proper level of trust (Helldin et al., 2013; Zhang et al., 2019).

Three main TiA layers have been highlighted (Hoff & Bashir, 2015; Marsh & Dibben, 2003). First, dispositional trust reflects the operator's stable and overall tendency to TiA and depends on factors such as age, gender, culture, and personality traits. Next, situational trust is influenced by current situational characteristics (e.g., workload or perceived risks) and the operator's contextual mental state (e.g., fatigue or mood).

Finally, learned trust is initially defined by prior experiences, beliefs, and knowledge, but it is dynamically updated during interaction, in relation to the automation's specific features and design. These three layers have an increasingly strong influence on the operator's reliance on the automation (Hoff & Bashir, 2015).

Automation's characteristics, which are linked mainly to learned trust, are known to influence human-robot trust (Hancock et al., 2011). Some of these elements have been studied in the context of TiAD, using driving simulators or Wizard of Oz vehicles. For example, an aggressive automated driving style seemed to decrease drivers' trust, while a lawful driving style increased it (Morris et al., 2017). This is consistent with previous results which noted that drivers most often preferred a defensive driving style (Strauch et al., 2019; Yusof et al., 2016). Nevertheless, other studies have found no primary effect of the driving style when comparing familiar (i.e., close to one's own manual style) vs. unfamiliar styles (Hartwich, Beggiato, et al., 2018) or familiar vs. defensive vs. dynamic driving styles (Beggiato et al., 2020). In some cases, automated driving had to anticipate and execute actions earlier than the driver would have done so (i.e., allowing a wider safe distance while passing another vehicle) to achieve the same declared level of trust (Abe et al., 2018). It was also found that manual driving inspired more trust on passengers' behalf than automated driving did (Strauch et al., 2019). These results were later confirmed, showing that a lack of familiarity with and knowledge of such systems decreased passengers' levels of TiAD, compared to human control (Schmidt et al., 2021). This finding may indicate that unexpected and sudden manoeuvres are likely to decrease drivers' TiAD. However, several studies have shown that takeover requests (TOR) did not decrease trust and increased drivers' understanding of the system. The authors postulated that TOR were, therefore, not considered as automation failures, but rather as normal automated driving features in these cases (Hergeth et al., 2016, 2017, 2015). Nevertheless, another study found that trust decreased after repeated TOR (Kraus, Scholz, Stiegemeier, et al., 2020), suggesting that such critical situations may have a negative impact on TiAD in some cases.

Regarding factors that influence drivers' dispositional and initial learned trust (Hoff & Bashir, 2015), studies showed that age (Hartwich, Beggiato, et al., 2018; Hartwich, Witzlack, et al., 2018) and personality traits (Kraus, Scholz, & Baumann, 2020) allowed for the prediction of TiAD in some situations. Initial information given about system performance may also have an impact on future trust construction (Kraus et al., 2019), because drivers' expectations and mental models are related to TiAD (Beggiato, Pereira, et al., 2015; Forster et al., 2019). Drivers' trust seems to increase with knowledge about system features (Khastgir et al., 2018) and familiarization with TOR (Hergeth et al., 2017). Promoting TiAD through positive textual or video information was also linked with longer glances towards a non-driving-related task (NDRT) compared to a lowered TiAD (Körber et al., 2018). This engagement in NDRT seems to be a reliable indication of drivers' TiAD and has been showed to correlate with their levels of driving automation (Beggiato, Hartwich, et al., 2015) and experience with the system (Forster et al., 2020). In addition, a correlation between lower self-reported trust and higher road monitoring have also been exhibited (Hergeth et al., 2016; Payre et al., 2017). Therefore, trust affects not only drivers' performance but also their visual strategies and non-driving related behaviour. Further, the initial level of TiAD seems to have an impact on trust construction and automation use during HAD (Beggiato, Hartwich, et al., 2015; Hartwich et al., 2020), which is the main focus of this paper.

The effect of the initial level of TiAD on drivers' behaviour among Trustful vs. Distrustful drivers was examined during the early use of a simulated automated driving system, specifically, during their first 30 minutes of experience with HAD. This allowed researchers to better understand driver's trust calibration processes at different moments during the early stages of HAD use and provided deeper insights into the influence of individual differences (i.e., the initial level of trust) in the driving context. Because HAD will likely be implemented first for highway use or in traffic jams (Wintersberger & Riener, 2016)–as has been confirmed by various studies (Becker & Axhausen, 2017; Kaur & Rampersad, 2018)–a monotonous highway scenario including two critical situations was used in this study. The HAD system in the present study did not trigger TOR during these situations, which furthers the understanding of drivers' trust construction and their engagement in spontaneous activities in such contexts. This research also investigated the visual strategies that are employed by drivers when they are not required to explicitly monitor the driving environment. Drivers' declared level of trust is known to increase over time when experiencing driving automation (Bueno et al., 2016; Gold et al., 2015; Hartwich, Witzlack, et al., 2018; Hergeth et al., 2016, 2017; Körber et al., 2018; Kraus, Scholz, Stiegemeier, et al., 2020). In the current experiment, it was hypothesized that drivers' level of trust would increase along with HAD use (H1). A low initial level of TiAD might be due to a poor mental model concerning HAD systems, which is likely to be readjusted when drivers experience simulated HAD (Beggiato & Krems, 2013). It was, therefore, hypothesized that the trust gain would be greater for Distrustful drivers than for Trustful drivers (H2) and that the differences between both groups would still be present at the end of the experiment (H3). Moreover, it has been argued that, for a correct trust calibration, operators need to experience system boundaries (Moray & Inagaki, 1999; Wintersberger et al., 2016). Considering the effects that TOR have on drivers' TiAD (Hergeth et al., 2016, 2017; Kraus, Scholz, Stiegemeier, et al., 2020), it was expected that the critical situations would lead to an abrupt increase of driving environment monitoring among drivers, which would then decrease after the event (H4). Trust questionnaires, gaze behaviour, and NDRA were analysed.

2. METHOD

2.1. Participants

Before the experimental session, the level of trust of 90 potential participants was assessed by email, using a dedicated scale (see Level of trust assessment). A *k-means* clustering method was then used to partition the potential participants in two categories, based on their initial level of trust. The 20 most extreme participants in each category were selected to create two experimental groups: Trustful drivers (those with the highest initial level of trust, n = 20, 6F/14M, M = 39.55 years old, SD = 9.09) and Distrustful drivers (those with the lowest initial level of trust, n = 20, 14F/6M, M = 36.50 years old, SD = 8.50). In total, 40 healthy adults (20 females, M = 38 years old, SD = 8.8) participated in a driving task in a single session. All participants had held a valid driver's license for a minimum of three years (M = 18.5, SD = 9.3) and drove regularly (M = 6.3 drives per week, SD = 1.6; M = 15,862 kilometres per year, SD = 12,822). Participants' perceived knowledge about HAD was assessed during the recruitment process on a four-point scale (no knowledge, basic knowledge, intermediate knowledge, or advanced knowledge). Among participants, 12.5% had no knowledge, 50% had a basic knowledge, and 37.5% had an intermediate knowledge. Participants were recruited by email and selected based on their declared initial level of TiAD (see Section 2.3). All drivers had normal or corrected-to-normal vision, and they received 40€ of compensation for their participation. This research complied with the American Psychological Association Code of Ethics and the European law on General Data Protection Regulation. Informed consent was obtained from each participant.

2.2. Apparatus

The study was conducted in a static driving simulator equipped with four 32'' 16/9 LCD screens giving a 120° horizontal field of view (Figure 1). The rear view was displayed on three digital mirrors, and dashboard information was provided by a 10” 16/9 LCD screen set behind the steering wheel. The driving simulation was controlled by the SCANeR™ Studio software, developed by AV Simulation, France (https://www.avsimulation.fr). A 10.1” Xenarc tablet installed in the central console of the simulator was used for HMI. A sideband on the left of the screen included a pictogram that informed drivers about the current vehicle state (i.e., “manual driving”, “available HAD” or “activated HAD”) as well as a button for HAD activation and specific pictograms when relevant. On the remaining portion of the screen, an Android™ emulator with games (e.g., Solitaire, Fruit Ninja, Mahjong) and internet access was displayed. The driver video-recording system was composed of four D-Link infrared cameras to investigate visual strategies and NDRA engagement.

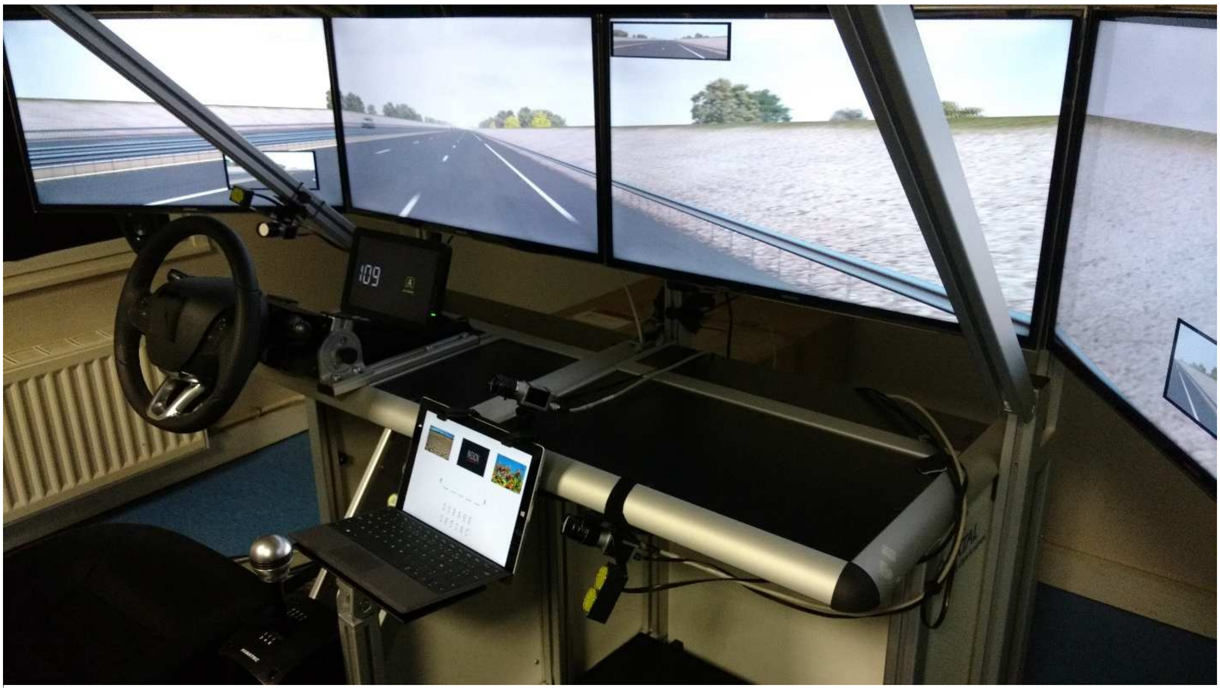

Figure 1. Driving simulator

2.3. Level of trust assessment

Initial and final level of learned TiAD were measured using a nine-item questionnaire (see Table 1) on a six-point Likert scale with responses ranging from 'not at all' to 'extremely' to assess trust and trust-related dimensions such as perceived safety, likelihood of use in degraded conditions, and utility. The scale was designed to specifically assess TiAD, in accordance with previous studies (see Table 1). A short paragraph described a level 4 automated driving system (SAE, 2016) right before participants answered the trust scale, to ensure all of them understood the same concept when referring to "automated vehicle". Items were formulated to be as inclusive as possible considering most participants had no prior experience with automated driving. In the final scale, items were formulated in the past tense, and were all directed towards the specific HAD system used during the experiment.

Table 1. Initial and final level of TiAD assessment scales

	Initial scale (before the experiment)	Final scale (after the experiment)	Origin
1	I would feel safe in an automated vehicle.	I felt safe in the automated vehicle.	O'Cass & Carlson (2012)
2	The automated driving system provides me with more safety compared to manual driving.	The automated driving system provided me with more safety compared to manual driving.	Payre et al. (2016)
3*	I would rather keep manual control of my vehicle than delegate it to the automated driving system on every occasion.	I would rather keep manual control of my vehicle than delegate it to the automated driving system on every occasion.	Payre et al. (2016)
4	I would trust the automated driving system decisions.	I trusted the automated driving system decisions.	O'Cass & Carlson (2012)
5	I would trust the automated driving system capacities to manage complex driving situations.	I trusted the automated driving system capacities to manage complex driving situations.	Egea & González (2011)
6	If the weather conditions were bad (e.g., fog, glare, rain), I would delegate the driving task to the automated driving system.	If the weather conditions were bad (e.g., fog, glare, rain), I would have delegated the driving task to the automated driving system.	Payre (2015)
7	Rather than monitoring the driving environment, I could focus on other activities confidently.	Rather than monitoring the driving environment, I could focus on other activities confidently.	Egea & González (2011)
8	If driving was boring for me, I would rather delegate it to the automated driving system than do it myself.	If driving was boring for me, I would rather delegate it to the automated driving system than do it myself.	Payre et al. (2016)
9	I would delegate the driving to the automated driving system if I was tired.	I would delegate the driving to the automated driving system if I was tired.	Payre et al. (2016)

*Answers were inverted for scoring.

Scale properties were analysed using *R* 3.6.1, (R Core Team, 2017) using the *psych* (Revelle, 2020) and *lavaan* packages (Rosseel, 2012). Homogeneity and internal consistency were good (Cronbach's α = .93, McDonald's ω_h = .78, and ω_t = .96). The factor structure of this nine-item scale was studied via exploratory factor analysis (EFA). Following the recommendations of Costello and Osborne (2005), a maximum likelihood factor analysis was employed using varimax rotation to study the expected single factor (i.e., trust). The results of the EFA showed that a one-dimensional factor accounted for 56.41% of the total variance of the data (Table 2). A confirmatory factor analysis (CFA) was then performed to better explore the scale's properties. Data were partitioned, with

60% used for training and the remaining 40% for testing. The testing yielded the following model: X^2 = 48.2, *df* = 27, *p* < .01, CFI = .80, RMSEA = .22 (CI low = .11; CI high = 0.32) and SRMR = .11 (Table 2). Following Brown's (2015) recommendations and given the small number of observations, the model seemed appropriate. The small number of observations also induced three items (Q3, Q6, and Q9) to have *p*-value higher than α = .05. These items have been kept in the trust scale, because previous works using higher number of observations showed they were relevant for trust assessment (Manchon et al., 2021).

	EFA			CFA			
Items	Loadings	Complexity	Uniqueness	Estimate	Standard Error	z-value	*p*
Q1	0.80	1.00	0.35	1.000			
Q2	0.86	1.00	0.26	1.853	0.846	2.189	0.029
Q3	0.55	1.00	0.69	0.734	0.545	1.348	0.178
Q4	0.77	1.00	0.40	1.852	0.871	2.126	0.034
Q5	0.74	1.00	0.45	2.539	1.130	2.247	0.025
Q6	0.82	1.00	0.32	1.144	0.700	1.635	0.102
Q7	0.88	1.00	0.22	2.088	0.947	2.204	0.028
Q8	0.69	1.00	0.52	2.216	1.065	2.081	0.037
Q9	0.54	1.00	0.71	1.659	0.858	1.932	0.053

Table 2. EFA and CFA loadings

Dynamic level of trust was measured using a Single Trust Item ('I trusted automated driving system decisions'; Item 4 from the previous scale, Table 1) on a six-point Likert scale to investigate current drivers' level of trust after each scenario (Lee & Moray, 1994; Seppelt & Lee, 2019).

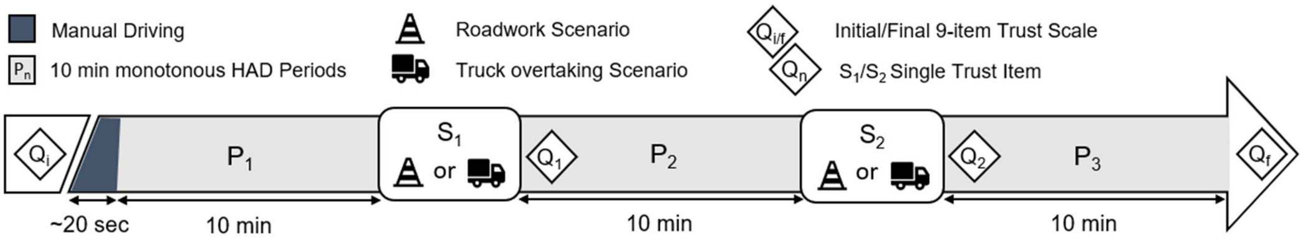

Figure 2. Experimental protocol

2.4. Procedure

Participants arrived at VEDECOM Institute and signed a consent form. They received a 10-minute practice to familiarize themselves with the driving simulator and the automated driving (e.g., HAD activation, vehicle manoeuvres during non-critical situations, pictograms, HMI, sounds). During the experimental task, participants activated HAD after merging onto the highway. The vehicle speed was set at 130 km/h– the maximum allowed speed on French highways–on a three-lane highway in low-density traffic (three to six vehicles per kilometre). Manual takeover was not possible, to ensure that all participants experienced the same scenarios, and they were free to engage in NDRA (e.g., reading, texting, using the Android™ tablet, listening to the radio). After 10 minutes of monotonous automated driving (P_1), participants were confronted with the first scenario (S_1), which was indicated by a sound and a specific pictogram five seconds prior to the event. They completed the first Single Trust Item immediately following this event (Q_1). The scenario was a roadwork area that was signalled by traffic cones and a roadwork van. The vehicle strongly decelerated from 130 km/h to 90 km/h, at about 6 m/s², until the time-to-collision (TTC) with the roadwork area was 2.5 seconds. It then changed from the right lane to the left lane and returned to the right lane after overtaking the obstacles. After the second 10-minute monotonous automated driving period (P_2), the second scenario (S_2) occurred, indicated by the same sound and a specific pictogram five seconds prior to the event, and followed by the second Single Trust Item (Q_2). This scenario was represented by a leading truck driving slowly (90 km/h) in the right lane, forcing the vehicle to decelerate from 130 km/h to 90 km/h at about 7 m/s^2, until the TTC with the truck was 1.5 seconds, because the left lane was congested with dense traffic. Once the left lane had been cleared, the vehicle overtook the truck and returned to the right lane. The order of presentation of the two scenarios was counterbalanced to neutralize potential order effects. Finally, the third 10-minute period of monotonous driving (P_3) and the final nine-item trust scale (Q_f) concluded the experimental session (Figure 2). Participants were then debriefed for approximately 10 minutes about their overall feelings concerning the experiment and their behaviour during HAD.

2.5. Dependent variables

Dependent variables included declared trust, glance count, glance duration, and NDRA engagement frequency.

Trust scales were scored averaging items, reversing item 3 (cf. Table 1). Item scores ranged from one to six. To compare overall trust evolution, the fourth item from the initial and final scales was used as a Single Trust Item.

Visual strategies and *NDRA* data were processed by manual video coding every one-tenth of one second. Due to technical problems, two Trustful participants were excluded from the analyses. For each change in gaze direction, one glance at the previously monitored area was counted (*glance count*), and the glance duration was added to the total glance duration towards that area, prior to being transformed into a percentage of the total glance time (*glance duration*). Areas of interest were defined as ‘road’, ‘rear mirrors’, ‘dashboard’, ‘NDRA’, ‘HMI sideband’, ‘Android™ tablet’, and ‘other’ for all glances directed elsewhere. ‘Road’, ‘rear mirrors’, ‘HMI sideband’, and ‘dashboard’ were grouped into the category of ‘driving environment’ for analysis. ‘NDRA’ and ‘Android™ tablet’ were grouped into the ‘NDRA’ category.

NDRA categories were defined based on participants’ most frequently observed activities: ‘mobile phone use’, ‘tablet use’, ‘reading’, and ‘radio use’. The other minor activities (e.g., grooming, drinking water, putting glasses on) were combined into a new category, ‘other’. These NDRA categories were opposed to ‘environment monitoring’, during which participants solely monitored the driving environment.

Data and graphics were processed using *R* 3.6.1 (R Core Team, 2017) and *ggplot2* (Wickham, 2016).

3. RESULTS

3.1. Declarative trust

The mean trust scores for the different questionnaires are reported in Table 3.

Table 3. Mean trust scores for the Single Trust Items.

Time	Group	Mean	SD	95% CI	
				Low	High
Initial Single Trust Item	Trustful	4.70	0.80	4.32	5.08
	Distrustful	2.90	0.79	2.53	3.27
1st Scenario	Trustful	4.65	1.27	4.06	5.24
	Distrustful	3.30	1.56	2.57	4.03
2nd Scenario	Trustful	5.10	1.17	4.55	5.65
	Distrustful	3.25	1.19	2.37	4.13
Final Single Trust Item	Trustful	5.30	0.73	4.96	5.64
	Distrustful	3.85	1.42	3.18	4.52

A two-way mixed-design ANOVA (Group x Time, 2 levels: Initial assessment and Final assessment) was performed to investigate the effect of the experience on trust construction in Trustful and Distrustful drivers (Figure 3). There was a significant effect of the Group, $F(1, 38) = 81.5$, $p < .001$, $\eta_p^2 = .682$, and the Time, $F(1, 38) = 12.3$, $p < .001$, $\eta_p^2 = .244$. The interaction Groupe x Time was also significant, $F(1, 38) = 8.21$, $p < .01$, $\eta_p^2 = .178$. In short, declared trust was found to increase for both Trustful and Distrustful participants during the interaction with the vehicle (H1), and the trust increase

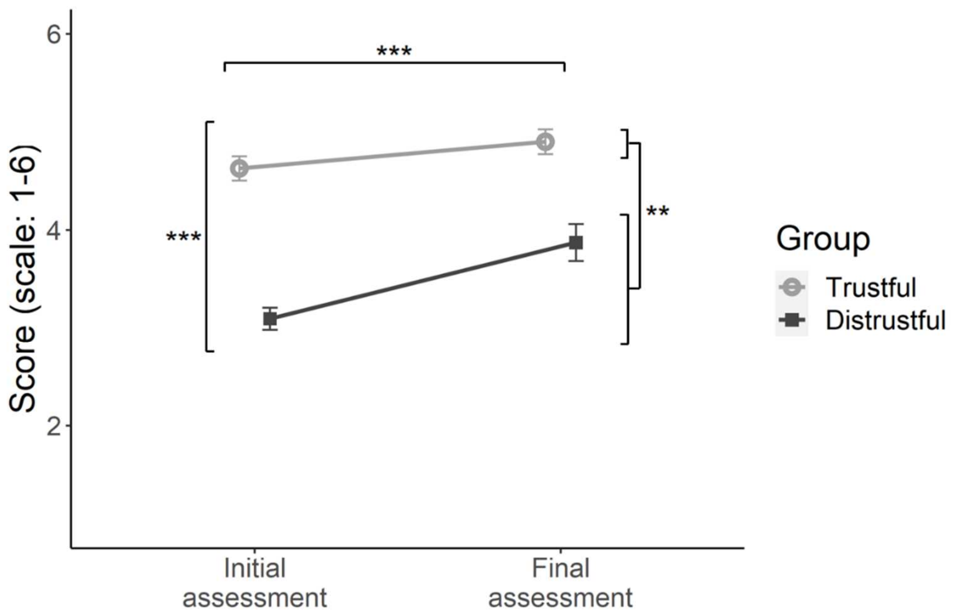

Figure 3. Initial and Final 9-item trust scales; error bars = standard error

was higher for Distrustful drivers (H2).

Another two-way mixed-design ANOVA (Group x Time, 4 levels: Initial assessment, 1st Scenario (S_1), 2nd Scenario (S_2) and Final assessment) was performed to investigate scenarios' impact on drivers' early trust construction (Figure 4). There was a significant effect of the Group, $F(1, 38) = 29.4$, $p < .001$, $\eta_p^2 = .436$, and the Time, $F(3, 114) = 4.66$, $p < .01$, $\eta_p^2 = .109$, but no interaction was found, $F(3, 114) = 0.65$, $p > .1$, $\eta_p^2 = .017$.

Post hoc tests using Bonferroni correction showed that the Initial assessment differed significantly from Final assessment ($p < .001$), but no other comparisons were significant. As expected, the differences between both groups were significant in the final assessment ($p < .001$), confirming H3.

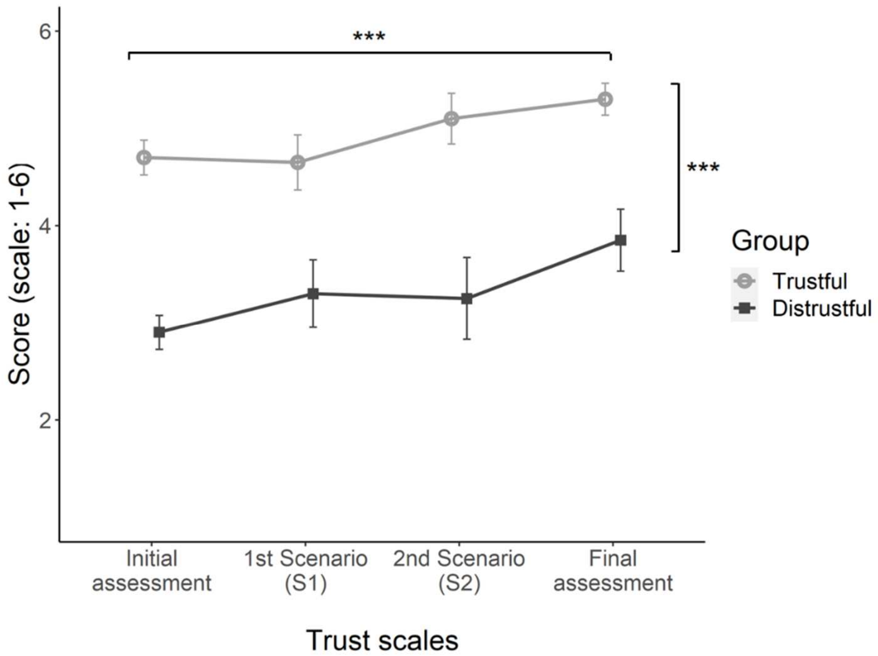

Figure 4. Declared level of TiAD evolution on Single Trust Items; error bars = standard error

Motivated by the differences in trust reported by participants concerning the two scenarios (62.5% of the participants declared *Truck* scenario as more critical, 20% *Roadwork* scenario, 17.5% did not feel any difference), a two-way mixed design ANOVA was conducted (Group x Scenario type, 2 levels: *Roadwork* scenario vs. *Truck* scenario) (Figure 5). There was a significant effect of the Scenario type, $F(1, 38) = 11.86$, $p < .001$, $\eta_p^2 = .238$ for both Trustful ($p < .001$) and Distrustful ($p < .001$) confirming that drivers reported less trust after the Truck scenario than after the Roadwork scenario.

Data were therefore separated to analyse whether this difference had an impact on early trust construction (Figure 6). A three-way mixed-design ANOVA (Group x Time x Scenario Order, 2 levels: *Roadwork → Truck* vs. *Truck → Roadwork*) showed a main effect of the Group, $F(1, 36) = 29.2$, $p < .001$, $\eta_p^2 = .448$, and Time, $F(3, 108) = 5.75$, $p < .001$, $\eta_p^2 = .138$. The main effect of Scenario Order was not significant, $F(1,$

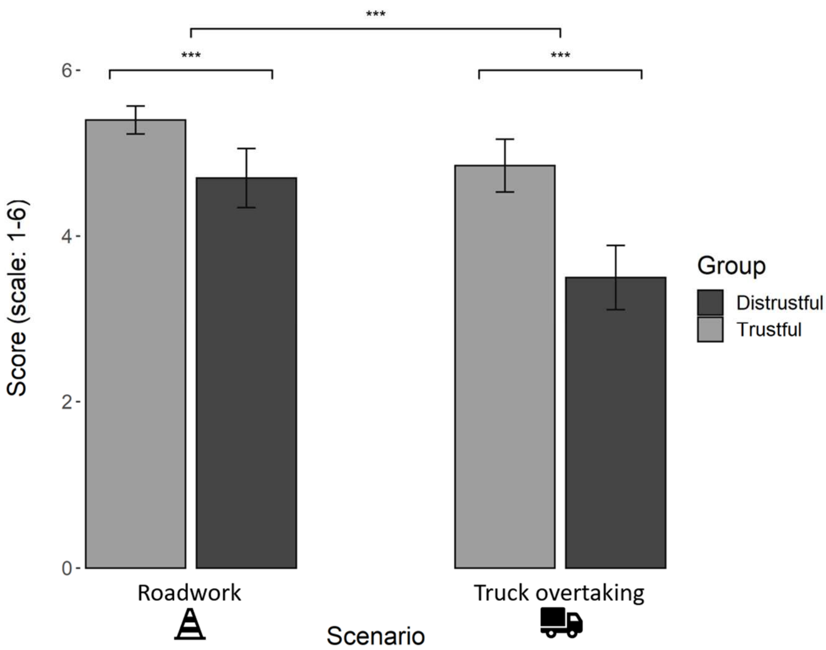

Figure 5. Declared level of TiAD depending on the Scenario; error bars = standard error

$36) = .044$, $p > .05$, $\eta_p^2 = .001$.

The interaction between Scenario Order and Time was significant, $F(3, 108) = 8.73$, $p < .001$, $\eta_p^2 = .195$. This revealed that, in sequence *Roadwork → Truck*, Initial assessment differed from S_1 ($p < .001$) and S_2 differed from Final assessment ($p < .001$), but in sequence *Truck → Roadwork*, Initial assessment differed from S_2

(p < .01). In short, early trust construction was found to differ according to Scenario Order for both Trustful and Distrustful participants.

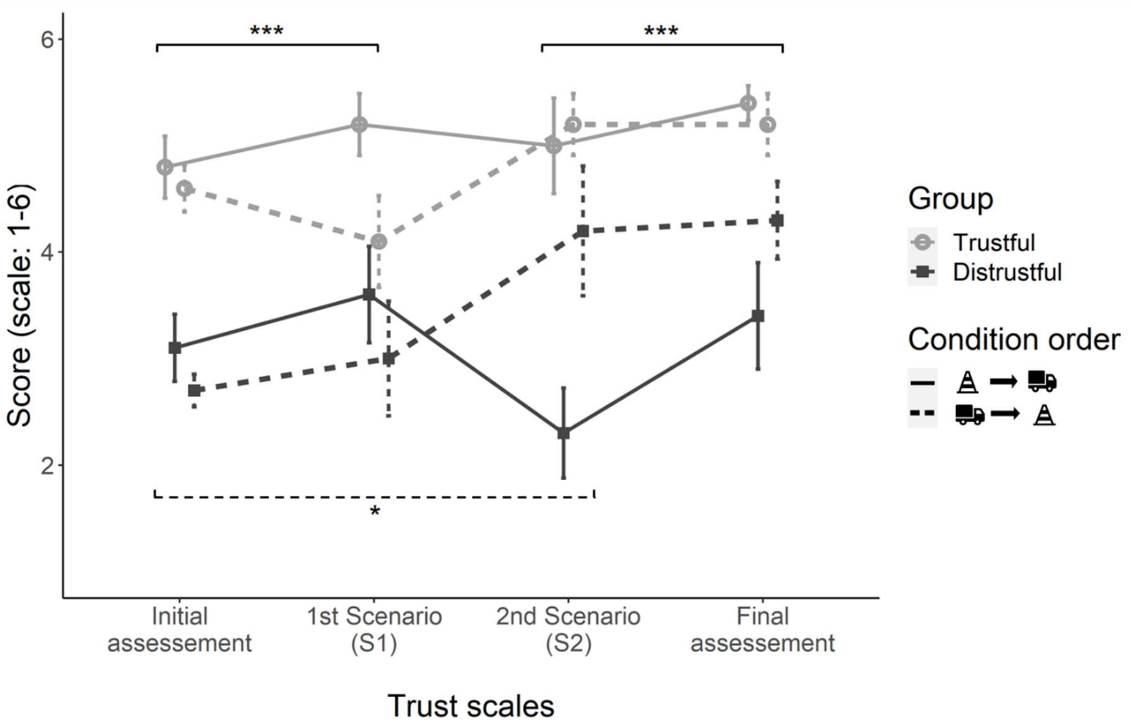

Figure 6. Declared level of TiAD evolution, depending on the Scenario Order; error bars = standard error

3.2. Visual behaviour

As stated in the Method section, two Trustful participants were excluded from the further analysis. The new Trustful group (n = 18) showed an initial level of trust such as M = 4.72, SD = 0.83. A t-test was run to compare the complete Trustful group (n = 20) and the smaller Trustful group (n = 18), $t(17) = 0$, $^2 = 1$, suggesting this reduction did not invalidate the nature of the Trustful group. A two-way mixed-design ANOVA (Group x Driving Period, 3 levels: P_1 vs. P_2 vs. P_3) was conducted on glance count and glance duration. Driving Periods were defined as the ten-minute period between the beginning of the experiment and the first critical scenario (P_1), the ten-minute period between both critical scenarios (P_2), and the remaining ten-minute period before the end of the scenario (P_3) (Figure 2). During monotonous driving, there was only a main effect of the Driving Period on glance count towards the Driving environment (Road, Rear-mirrors

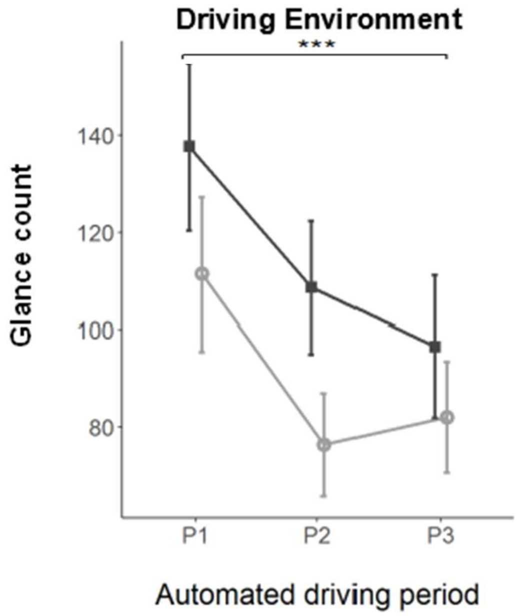

Figure 7. Glance count towards the driving environment during the automated Driving Periods; error bars = standard error

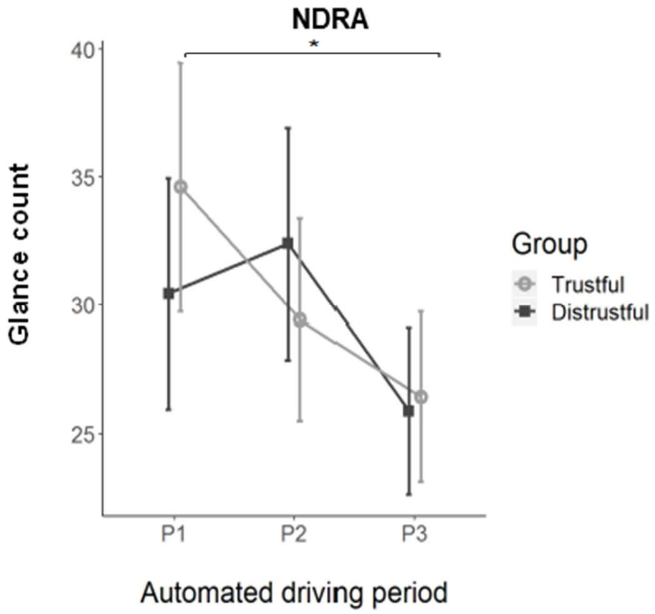

Figure 8. Glance count towards the NDRA during the automated Driving Periods; error bars = standard error

and Dashboard), $F(2, 72) = 18.6$, $p < .001$, $\eta_p^2 = .341$ (Figure 7), and towards the NDRA, $F(2, 72) = 3.54$, $p < .05$, $\eta_p^2 = .090$ (Figure 8). A post hoc test showed that glance count towards the Driving environment was reduced between P_1 and P_2 ($p < .001$), but no difference was found between P_2 and P_3 ($p > .1$). In short, drivers' number of eyes movements decreased during HAD for both groups, and this reduction mainly concerned the first 10 minutes of HAD use.

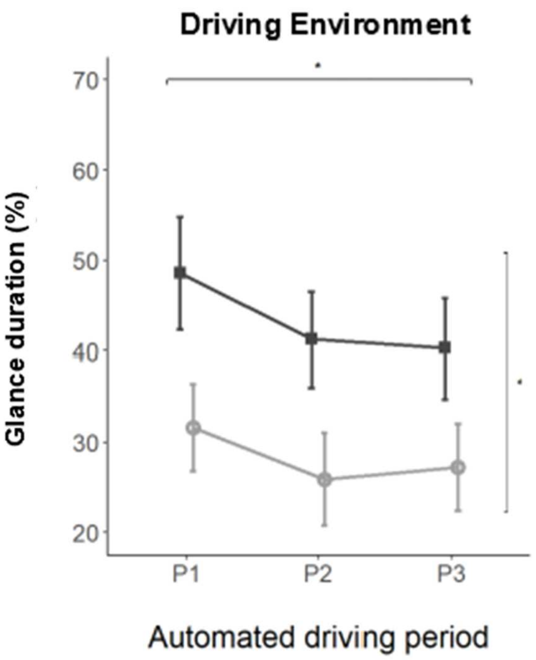

Figure 9. Proportion of glance duration (%) towards the driving environment during the automated Driving Periods; error bars = standard error

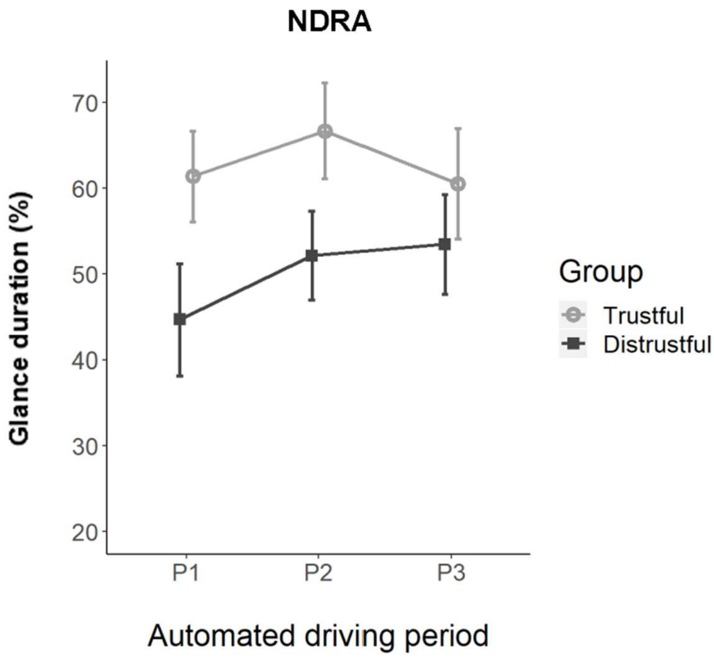

Figure 10. Proportion of glance duration (%) towards NDRA during the automated Driving Periods; error bars = standard error

Concerning glance duration towards the Driving environment, results showed a main effect of the Group, $F(1, 36) = 4.64$, $p < .05$, $\eta_p^2 = .114$ and a main effect the Driving Period, $F(2, 72) = 4.28$, $p < .05$, $\eta_p^2 = .106$ (Figure 9). A post hoc test showed that glance duration was reduced between P_1 and P_2 ($p < .05$), but no difference was found between P_2 and P_3 ($p > .1$). There were no significant differences concerning glance duration towards NDRA (Figure 10). In short, drivers monitored road less frequently and during less time (H1) over HAD and particularly during the first 10 minutes. Distrustful drivers tent to check the driving environment during more time than Trustful drivers (H3).

In order to evaluate the visual behaviour evolution over time, a binary logistic regression with repeated measures (*logit*) opposing glances towards the driving environment (Road, Mirrors, and Dashboard) to all glances directed elsewhere (NDRA and Other) was conducted. Temporal variables were time (squared) and 3 variables constructed from simple, squared, and cubed values of the distance to each of the Scenarios (S_1 & S_2).

The binary logistic regression allows to bind a probability (*p*, here the probability that a particular glance is directed towards the driving environment) to the glance's characteristics (X_i) and determines the intensity of these bonds (β_i). It relies on a formula ($Ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \cdots + \beta_i * X_i$) were *p* corresponds to the probability the driver is looking towards the driving environment, X_i correspond to the glance's characteristics (e.g., is it from a male or female, is it from a Trustful or a Distrustful, how many Time passed since the beginning of the experiment), and β_i correspond to the model's estimate parameters.

The final model (Table 4, Figure 11) includes all the variables described above. The *logit* was controlled with Group, Gender and Conditions order. Reference variables are indicated in italic. Due to the high number of observations in the model, the *p*-values may be higher than they would be with ANOVAs. The dynamic of the variation, on the other hand, can be fully considered.

Table 4. Influence of specific periods of time on glances directed towards the driving environment between Trustful and Distrustful drivers

	Trustful	
	β	Odds ratio
(Intercept)	4.672***	106.963
Female	-2.120*	0.120
Distrustful	2.080*	8.008
Time (^{2})	-0.113***	0.893
Interaction Time (^{2}) x Group	0.007***	1.007
Condition	-0.937	0.391
Distance to Scenario		

Distance to S_1	1.348***	3.849
Distance to S_1 (^{2})	-2.152***	0.116
Distance to S_1 (^{3})	1.098***	2.999
Distance to S_2	1.364***	3.911
Distance to S_2 (^{2})	-3.533***	0.029
Distance to S_2 (^{3})	2.509***	12.301
AIC	*521415.4*	
ROC	*0.669*	
Number of observations	*564787*	

The results of this model (Table 4, Figure 11) showed again that the probability to look towards the driving environment during critical scenarios was higher for Distrustful drivers than for Trustful drivers. Moreover, the probability to look towards the driving environment was higher immediately after the scenarios than during the next driving

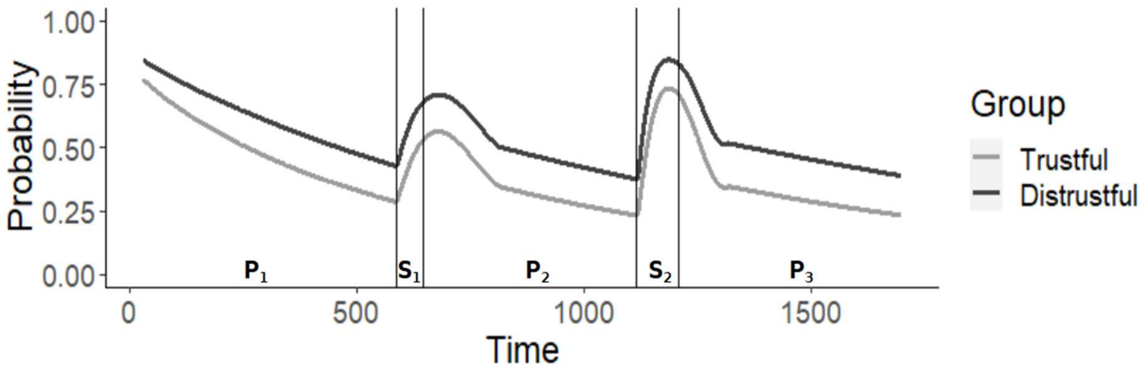

Figure 11. Evolution of the probability to look towards the driving environment over time.

period (for both drivers). These results confirmed the H4.

3.3. Non-driving-related activities

For the further analysis, two Trustful participants (the same that were excluded in the previous analysis) were not included. Two participants (1 Trustful and 1 Distrustful) did not engage in NDRA at all. During HAD, participants monitored the driving environment 21.9% of their time. The rest of the time, the most frequent NDRA they engaged in were mobile use for texting, calling, playing or web browsing (33.6%), tablet use for playing games and web browsing (17%), reading magazines or other documents (14.2%), and listening to the radio (3.75%). Other minor activities represented 9.15% of participants' time (Figure 12). A chi-square test of independence between Group and NDRA engagement was significant, $\chi^2(5, N = 38) = 27.272$, $p < .001$. Distrustful drivers were more likely to monitor the driving environment and less likely to engage in some NDRA

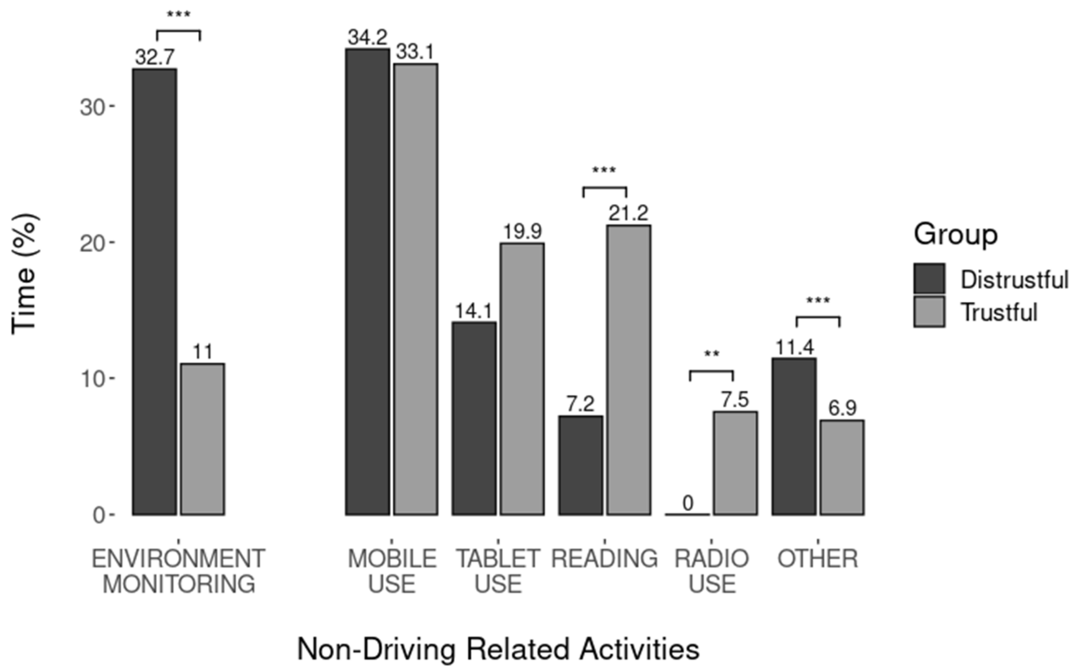

Figure 12. Drivers' environment monitoring and NDRA engagement during HAD.

(i.e., reading, radio use) than were Trustful drivers (H2).

4. Discussion

The current study aimed to investigate the impact of simulated HAD experience (an overall level of trust increase was expected, H1), the influence of drivers' initial level of TiAD (the trust gain was expected to be greater for Distrustful drivers than Trustful drivers, H2; the difference in trust between both groups was expected to exist at the end of the experiment, H3), and the effects of safety-critical situations on drivers' visual strategies (which were expected to increase their driving environment monitoring during the critical situations, H4) during the first interactions with automated driving. Drivers' early trust construction during these 30 minutes drives was examined through questionnaires, visual behaviours, and non-driving related activities.

4.1. Impact of HAD experience

Reported trust was found to increase over time, particularly during the first 10 minutes of the experiment; this finding supports H1. This is consistent with previous results that showed an overall positive effect of the simulated HAD on trust (Gold et al., 2015; Hergeth et al., 2016). Similarly, Hartwich et al. (2018) observed a strong increase in trust during the first interaction with the system, followed by a stabilization. Glance behaviours are known to be influenced by driving automation introduction (e.g., Navarro et al., 2016, 2017, 2019). It was found in this experiment that participants' glance behaviour changed during the driving session and did so in a similar way for both groups. Glance count towards the driving environment decreased by 28%, while glance count towards NDRA decreased by 20%. This overall decreasing trend in glance count, combined with a decrease in glance duration towards the driving environment, indicated a change in drivers' visual strategy. Participants seemed to monitor the road less and experienced increasing periods of fixed gaze. Moreover, results showed the decrease in glance count towards the driving environment mainly took place between P_1 and P_2, suggesting the decrease in road monitoring appeared rapidly after the beginning of the interaction. Contrarily, the glance count towards the NDRA decreased between P_2 and P_3, and not before. This result suggests drivers had fewer, but longer glances towards the NDRA after about 20 minutes of interaction with HAD. This may indicate an increase in trust but may also be the result of a driver's habituation to the driving simulation. Nevertheless, these data are consistent with previous results (Hergeth et al., 2016) and statements regarding trust influence on automation monitoring (e.g., Muir & Moray, 1996). Given that drivers' low level of trust is likely to lead to a spontaneous manual recovery (Payre, Cestac, & Delhomme, 2016), this information raises concerns about drivers' ability to retake manual control in urgent situations if their mental representation of the surrounding environment is not current. Such manoeuvres could be more dangerous than HAD. The provision of specific feedback indicating that the HAD system is performing normally (Beller et al., 2013; Helldin et al., 2013; Wintersberger et al., 2019) could limit drivers' urgent need to take control.

4.2. Impact of initial level of TiAD

Reported trust increased for both groups. However, Distrustful participants' trust rose significantly more after they had experienced HAD, as showed by the 9-item scale (0.08 points for Trustful drivers, Cohen's *d* = .15 and 0.77 points for Distrustful drivers, Cohen's *d* = .98, on the 1 to 6 Likert-type scale), thereby supporting H2. Nevertheless, the Initial-Final trust comparison using the Single Trust Item showed no such interaction in trust progression. This result may indicate the Single Trust Item is less sensible than the 9-item scale, which integrate more information about participants' level of trust. This difference between trust scales measures should be examined in future studies. However, participants declared differences in HAD perception, expectation, and use intention. Thus, most of the Distrustful drivers felt uncomfortable with the idea of using automated driving on the road, while Trustful drivers were enthusiastic about the NDRA engagement possibilities. The initial level of trust may therefore be used as a good indicator of HAD use and driver's behaviour in the early use of automated driving. Moreover, Trustful and Distrustful passengers seem to have divergent needs concerning the continuity of feedback: Distrustful passengers may require continuous information about the HAD performance, while Trustful passengers may prefer situation-specific information only (Hartwich et al., 2021). More research is needed regarding drivers' needs for safer and more pleasant feedback from the vehicle relative to their initial levels of trust. On one hand, Trustful drivers may opt for less feedback in order to reduce their cognitive load while they engage in NDRA. On the other hand, Distrustful drivers may benefit from continuous head-up display feedback that provides them knowledge about HAD performance without interrupting their monitoring of the road.

In addition, reported trust was found to be consistently lower for Distrustful drivers than for Trustful drivers at each of the four assessments of trust made during the experiment; this finding supports H3. This result could indicate that pre-existing knowledge or beliefs influencing initial learned trust (Hoff & Bashir, 2015) can have a persistent effect on drivers' early trust construction during HAD. It, therefore, seems that 30 minutes of HAD is not enough to erase the difference in levels of trust between Distrustful and Trustful drivers. A longer HAD experience may be needed to overcome this initial low level of trust, as has been shown in longer studies in other human-machine interaction contexts (Mayer, 2008; Sauer et al., 2015).

Drivers' visual strategies were consistent with declared level of trust: Distrustful drivers spent 43.4% of their time monitoring the road, in contrast to only 28.3% by Trustful drivers. This confirms previous findings by Hergeth et al. (2016). Accordingly, time allocated to NDRA showed the inverse pattern–50% of the time among Distrustful drivers vs. 62.8% among Trustful drivers–which is consistent with the results attained by Körber et al. (2018). Drivers' NDRA engagement confirmed the influence of the initial level of TiAD, as Distrustful drivers engaged less in some activities (e.g., reading, using the radio) and were more likely to monitor the driving environment. This finding contributes new elements that support the understanding of initial learned trust influence on early trust construction (Hoff & Bashir, 2015) in the context of HAD. Finally, the regression model showed an impact of gender on gaze behaviours. This effect was not found in other analyses (i.e., scores and NDRA engagement) nor in other studies (Feldhütter et al., 2016; Molnar et al., 2018; Schwarz et al., 2019). However, the imbalance in female/male distribution in both groups may have had an impact on these results, and further studies are needed to clarify the link between gender and initial level of trust.

4.3. Impact of critical situations

The two scenarios presented in this experiment were realistic and common driving situations. Both situations induced a small trust drop and increased temporarily drivers' glances towards the driving environment, a finding that supports H4. Although the scenarios were designed to be similar, results indicated they were perceived differently.

Considering post-experimental debriefings, this difference is likely due to the roadwork scenario being perceived as safer than the truck overtaking scenario because the first one implied a simple lateral manoeuvre while the second one required both a longitudinal and a lateral manoeuvre. Furthermore, the roadwork area scenario was free of any traffic or pedestrian, while the truck overtaking scenario included several other cars and the truck. Therefore, dissociating both scenarios provided information about how perceived safety may impact early trust construction. The results suggested that trust evolves differently depending on exposure to critical situations. If drivers are exposed to a more critical situation at the very first interaction with automated driving, trust will not increase, but will remain stable; it will then grow gradually after a situation that is perceived as less dangerous. This may indicate that drivers are calibrating their trust and, therefore, are carefully observing the next AV action. Conversely, if the first interaction is perceived as safer than the second, trust is likely to increase significantly–potentially leading to overtrust–but also to return to its initial level after the second event. This supports the results of Walker et al. (2019) regarding the importance of early interaction in TiAD construction. Trust calibration seems to be more irregular. Again, this reduction in TiAD may lead to a manual recovery (Payre, Cestac, & Delhomme, 2016), possibly resulting in a poorly realized manoeuvre.

These results indicate that first experiences with automated driving systems may have a stronger influence on short-term trust calibration. Proper drivers' training for automated driving is a current issue in human factors research (Payre et al., 2016; Wintersberger et al., 2016). Our results suggest that it may be valuable to experience a critical situation during the first interactions with automated driving, particularly for Distrustful people, in order to improve drivers' trust calibration. Nonetheless, given the small number of participants in each condition, this finding needs to be confirmed by further research.

4.4. Limitations and perspectives

Because of the study's design, participants were not randomly assigned to either the Trustful or the Distrustful groups in the current study but were selected prior to the experiment, based on their initial trust score. This *quasi-experimental* approach was chosen to ensure that both groups were homogenous and contained particularly Trustful or particularly Distrustful people, in contrast to a posterior median-split design. Because participants had various profiles and no common points, except for their initial level of trust, this factor does not seem to pose a threat to the study's internal validity. Nevertheless, this must be taken into account when the study's results are examined.

In this experiment, to guarantee that all participants experienced both scenarios, takeover control of the vehicle was not allowed. This methodological choice increased experimental control but was less natural, as participants could not return to manual control. Other studies are also required in order to investigate the influence of the timing of critical situations during one or more sessions. As has been stated by Hoff and Bashir (2015), system malfunctions or operators' experience with the system may impact learned trust. Here, each condition occurred after 10 minutes of HAD in a single session; however, trust might evolve differently after longer HAD periods of use, multiple driving sessions, or a higher number of critical scenarios experienced during HAD. Furthermore, it may be valuable to investigate the impact of other trust factors (Hoff & Bashir, 2015), such as different driving environments or mental workload, on drivers' behaviour during HAD use. Moreover, the simulated environment provided a widely replicable experiment, but it also decreased participants' perceived risk. On-road studies may explore TiAD related factors with a less-biased feeling of safety by participants.

5. Conclusion

This study confirms previously established relationships between self-reported trust and road monitoring during HAD and offers insights into drivers' possible NDRA engagement following HAD. It provides additional information regarding the influence of the initial level of trust on further trust development during the first interactions with HAD. It also shows that drivers with high initial levels of TiAD are more likely to engage in NDRA, in contrast with drivers with low initial levels of TiAD. These initial levels of trust may influence the type of NDRA that drivers engage in, as Trustful drivers seems to be more prone to read than Distrustful ones.

Car manufacturers should be aware of these effects and may use simulated HAD to help drivers calibrate an appropriate level of TiAD relative to the capabilities of a particular HAD system.

ACKNOWLEDGMENTS

We wish to thank Romane Beaufort for her advice on statistical analyses, and five anonymous reviewers for their helpful comments on a previous version of this work.

Funding: This study was supported by VEDECOM Institute, ‘‘Institute for Energy Transition” and part of the French governmental plan ‘‘Investment for the Future” (ANR-10-IEED-0009).

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6. REFERENCES

Abe, G., Sato, K., & Itoh, M. (2018). Driver trust in automated driving systems: The case of overtaking and passing. *IEEE Transactions on Human-Machine Systems*, *48*(1), 85–94. https://doi.org/10.1109/THMS.2017.2781619

Beggiato, M., Hartwich, F., Roßner, P., Dettmann, A., Enhuber, S., Pech, T., Gesmann-Nuissl, D., Mößner, K., Bullinger, A. C., & Krems, J. (2020). KomfoPilot—Comfortable Automated Driving. In G. Meixner (Ed.), *Smart Automotive Mobility: Reliable Technology for the Mobile Human* (pp. 71–154). Springer International Publishing. https://doi.org/10.1007/978-3-030-45131-8_2

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.* https://doi.org/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. https://doi.org/10.1016/j.trf.2012.12.006

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

Beller, J., Heesen, M., & Vollrath, M. (2013). Improving the driver–automation interaction: An approach using automation uncertainty. *Human Factors: The Journal of the Human*

Factors and Ergonomics Society, *55*(6), 1130–1141. https://doi.org/10.1177/0018720813482327

Brown, T. A. (2015). *Confirmatory factor analysis for applied research, 2nd ed* (pp. xvii, 462). The Guilford Press.

Bueno, M., Dogan, E., Selem, F. H., Monacelli, E., Boverie, S., & Guillaume, A. (2016). How different mental workload levels affect the take-over control after automated driving. *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference On*, 2040–2045.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *Vol. 10.*, Article 7. https://doi.org/10.7275/JYJ1-4868

Feldhütter, A., Gold, C., Hüger, A., & Bengler, K. (2016). Trust in Automation as a matter of media influence and experience of automated vehicles. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 2024–2028. https://doi.org/10.1177/1541931213601460

Forster, Y., Geisel, V., Hergeth, S., Naujoks, F., & Keinath, A. (2020). Engagement in non-driving related tasks as a non-intrusive measure for mode awareness: A simulator study. *Information*, *11*(5), 239. https://doi.org/10.3390/info11050239

Forster, Y., Hergeth, S., Naujoks, F., Beggiato, M., Krems, J. F., & Keinath, A. (2019). Learning and development of mental models during interactions with driving automation: A simulator study. *Proceedings of the 10th International Driving Symposium on Human*

Factors in Driver Assessment, Training and Vehicle Design: Driving Assessment 2019, 398–404. https://doi.org/10.17077/drivingassessment.1724

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

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*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254

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. https://doi.org/10.1080/00140139.2018.1441448

Hartwich, F., Hollander, C., Johannmeyer, D., & Krems, J. F. (2021). Improving passenger experience and Trust in Automated Vehicles through user-adaptive HMIs: "The more the better" does not apply to everyone. *Frontiers in Human Dynamics*, *3*, 669030. https://doi.org/10.3389/fhumd.2021.669030

Hartwich, F., Schmidt, C., Gräfing, D., & Krems, J. F. (2020). In the passenger seat: Differences in the perception of human vs. Automated Vehicle control and resulting HMI demands of users. In H. Krömker (Ed.), *HCI in Mobility, Transport, and Automotive Systems. Automated Driving and In-Vehicle Experience Design* (Vol. 12212, pp. 31–45). Springer International Publishing. https://doi.org/10.1007/978-3-030-50523-3_3

Hartwich, F., Witzlack, C., Beggiato, M., & Krems, J. F. (2018). 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*. https://doi.org/10.1016/j.trf.2018.05.012

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 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '13*, 210–217. https://doi.org/10.1145/2516540.2516554

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

Hergeth, S., Lorenz, L., Krems, J., & Toenert, L. (2015). *Effects of take-over requests and cultural background on automation trust in highly automated driving*. 331–337. https://doi.org/10.17077/drivingassessment.1591

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

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

Khastgir, S., Birrell, S., Dhadyalla, G., & Jennings, P. (2018). Calibrating trust through knowledge: Introducing the concept of informed safety for automation in vehicles. *Transportation Research Part C: Emerging Technologies*, *96*, 290–303. https://doi.org/10.1016/j.trc.2018.07.001

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

Kraus, J., Forster, Y., Hergeth, S., & Baumann, M. (2019). Two routes to trust calibration: Effects of reliability and brand information on Trust in Automation. *International Journal of Mobile Human Computer Interaction*, *11*(3), 17.

Kraus, J., Scholz, D., & Baumann, M. (2020). What's driving me? Exploration and validation of a hierarchical personality model for trust in automated driving. *Human Factors*, 30. https://doi.org/10.1177/0018720820922653

Kraus, J., Scholz, D., Stiegemeier, D., & 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: The Journal of the Human Factors and Ergonomics Society*, *62*(5), 718–736. https://doi.org/10.1177/0018720819853686

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

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

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

Manchon, J. B., Bueno, M., & Navarro, J. (2020). From manual to automated driving: How does trust evolve? *Theoretical Issues in Ergonomics Science*, 1–27. https://doi.org/10.1080/1463922X.2020.1830450

Manchon, J. B., Bueno, M., & Navarro, J. (2021). Calibration of Trust in Automated Driving: A matter of initial level of trust and automated driving style? *Human Factors*, *(accepted)*.

Marsh, S., & Dibben, M. R. (2003). The role of trust in information science and technology. *Annual Review of Information Science and Technology*, *37*(1), 465–498. https://doi.org/10.1002/aris.1440370111

Mayer, A. K. (2008). *The manipulation of user expectancies: Effects on reliance, compliance, and trust using an automated system*. Georgia Institute of Technology.

Molnar, L. J., Ryan, L. H., Pradhan, A. K., Eby, D. W., St. Louis, R. M., & 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*, 319–328. https://doi.org/10.1016/j.trf.2018.06.004

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. https://doi.org/10.1177/014233129902100408

Morris, D. M., Erno, J. M., & Pilcher, J. J. (2017). Electrodermal response and automation trust during simulated self-driving car use. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 1759–1762. https://doi.org/10.1177/1541931213601921

National Highway Traffic Safety Administration, , Department of Transportation. (2017). *INCLA-PE16-007*. https://static.nhtsa.gov/odi/inv/2016/INCLA-PE16007-7876.PDF

Navarro, J. (2019). A state of science on highly automated driving. *Theoretical Issues in Ergonomics Science*, *20*(3), 366–396. https://doi.org/10.1080/1463922X.2018.1439544

Navarro, J., François, M., & Mars, F. (2016). Obstacle avoidance under automated steering: Impact on driving and gaze behaviours. *Transportation Research Part F: Traffic Psychology and Behaviour*, *43*, 315–324. https://doi.org/10.1016/j.trf.2016.09.007

Navarro, J., Osiurak, F., Ovigue, M., Charrier, L., & Reynaud, E. (2019). Highly automated driving impact on drivers' gaze behaviors during a car-following task. *International Journal of Human–Computer Interaction*, 1–10. https://doi.org/10.1080/10447318.2018.1561788

Navarro, J., Reynaud, E., & Gabaude, C. (2017). Eye movement analysis in dynamic scenes: Presentation and application of different methods in bend taking during car driving. *Le Travail Humain*, *80*(3), 307. https://doi.org/10.3917/th.803.0307

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

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

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

R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2020). *psych: Procedures for personality and psychological research* (R package version 2.0.8) [Computer software]. North-western University, Evanston.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

SAE International. (2016). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. https://doi.org/10.4271/J3016_201609

Sauer, J., Chavaillaz, A., & Wastell, D. (2015). Experience of automation failures in training: Effects on trust, automation bias, complacency and performance. *Ergonomics*, *59*(6), 767–780. https://doi.org/10.1080/00140139.2015.1094577

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

Schmidt, C., Hartwich, F., & Krems, J. F. (2021). Looking at driving automation from a passenger's perspective: Driving experience and gaze behavior during fully automated vs. human vehicle control. In T. Ahram, R. Taiar, K. Langlois, & A. Choplin (Eds.), *Human Interaction, Emerging Technologies and Future Applications III* (Vol. 1253, pp. 3–9). Springer International Publishing. https://doi.org/10.1007/978-3-030-55307-4_1

Schwarz, C., Gaspar, J., & Brown, T. (2019). The effect of reliability on drivers' trust and behavior in conditional automation. *Cognition, Technology & Work*, *21*(1), 41–54. https://doi.org/10.1007/s10111-018-0522-y

Seppelt, B. D., & 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*, 66–80. https://doi.org/10.1016/j.ijhcs.2018.12.009

Strauch, C., Mühl, K., Patro, K., Grabmaier, C., Reithinger, S., Baumann, M., & Huckauf, A. (2019). *Real autonomous driving from a passenger's perspective: Two experimental investigations using gaze behaviour and trust ratings in field and simulator*. 14.

Walker, F., Wang, J., Martens, M. H., & 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*, 401–412. https://doi.org/10.1016/j.trf.2019.05.021

Wintersberger, P., Dmitrenko, D., Schartmüller, C., Frison, A.-K., Maggioni, E., Obrist, M., & Riener, A. (2019). S(C)ENTINEL: Monitoring automated vehicles with olfactory reliability displays. *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 538–546. https://doi.org/10.1145/3301275.3302332

Wintersberger, P., Frison, A.-K., Riener, A., & Boyle, L. N. (2016). Towards a personalized trust model for highly automated driving. *Proceeding of M&C*, *4*, 8.

Wintersberger, P., & Riener, A. (2016). Trust in technology as a safety aspect in highly automated driving. *I-Com*, *15*(3). https://doi.org/10.1515/icom-2016-0034

Yusof, N. Md., Karjanto, J., Terken, J., Delbressine, F., Hassan, M. Z., & Rauterberg, M. (2016). The Exploration of Autonomous Vehicle Driving Styles: Preferred Longitudinal, Lateral, and Vertical Accelerations. *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - Automotive'UI 16*, 245–252. https://doi.org/10.1145/3003715.3005455

Zhang, T., Tao, D., Qu, X., Zhang, X., Lin, R., & Zhang, W. (2019). The roles of initial trust and perceived risk in public's acceptance of automated vehicles. *Transportation Research Part C: Emerging Technologies*, *98*, 207–220. https://doi.org/10.1016/j.trc.2018.11.018