Transportation Research Part F 64 (2019) 401–412

Contents lists available at ScienceDirect

Transportation Research Part F

journal homepage: www.elsevier.com/locate/trf

Gaze behaviour and electrodermal activity: Objective measures of drivers' trust in automated vehicles

F. Walker [a,*], J. Wang [a], M.H. Martens [a,b], W.B. Verwey [a]

[a] University of Twente, Drienerlolaan 5, 7522 NB Enschede, the Netherlands
[b] TNO Human Factors, Kampweg 5, 3769 DE Soesterberg, the Netherlands

ARTICLE INFO

Article history:
Received 29 September 2018
Received in revised form 30 May 2019
Accepted 30 May 2019

Keywords:
Trust in automation
Trust calibration
Eye movement behaviour
Electrodermal Activity
Automated driving
Secondary task

ABSTRACT

Studies show that drivers' intention to use automated vehicles is strongly modulated by trust. It follows that their benefits are unlikely to be achieved if users do not trust them. To date, most studies of trust in automated vehicles have relied on self-reports. However, questionnaires cannot capture real-time changes in drivers' trust, and are hard to use in applied settings. In previous work, we found evidence that gaze behaviour could provide an effective measure of trust. In this study we tested whether combining gaze behaviour with Electrodermal Activity could provide a stronger metric. The results indicated a strong relationship between self-reported trust, monitoring behaviour and Electrodermal Activity: The higher participants' self-reported trust, the less they monitored the road, the more attention they paid to a non-driving related secondary task, and the lower their Electrodermal Activity. We also found evidence that combined measures of gaze behaviour and Electrodermal Activity predict self-reported trust better than either of these measures on its own. These findings suggest that such combined measures have the potential to provide a reliable and objective real-time indicator of driver trust.

1. Introduction

Automated vehicles promise to shape the future of mobility, for drivers and society as a whole. They promise to increase drivers' comfort by exempting them from routine driving tasks and by giving them the opportunity to perform secondary activities while the car takes care of driving (Gold, Körber, Hohenberger, Lechner, & Bengler, 2015). Drivers who have lost the ability to drive due to age or disability will regain back their mobility. Furthermore, automated vehicles are expected to drastically reduce road accidents and improve traffic flow, while reducing fuel consumption and emissions (Gold et al., 2015; Payre, Cestac, & Delhomme, 2014; Urmson & Whittaker, 2008). However, these benefits will only be achieved if drivers actually adopt the technology and this requires trust (Choi & Ji, 2015; Bailey & Scerbo, 2007; Hoff & Bashir, 2015; Lee & See, 2004).

Trust has been widely investigated in the field of interpersonal relationships (e.g., Rotter, 1980; Rempel, Holmes, & Zanna, 1985; Guenzi & Georges, 2010). Nevertheless, in the automotive domain, the most accepted definition is the one provided by Lee and See (2004). Those authors define trust as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (p. 51). Trust is thus an attitude, in which characteristics and qualities of the agent – in this case the automated vehicle - play a crucial role (Hoff & Bashir, 2015). Importantly, trust does not only

* Corresponding author.
E-mail address: f.walker@utwente.nl (F. Walker).

https://doi.org/10.1016/j.trf.2019.05.021
1369-8478/

predict whether the automated system will be used, but also *how* it will be used. Low trust can lead to disuse while, conversely, high trust can prompt abuse (Lee & Moray, 1994; Parasuraman & Riley, 1997; Payre et al., 2014). Disuse and abuse of automation can be seen as extremes on a single scale. Drivers' position on this scale and their acceptance of automated technology depends on their initial trust, but also on the vehicle's perceived reliability, and on the drivers' experience with the automated system (Walker, Boelhouwer, Alkim, Verwey, & Martens, 2018; Körber, Baseler, & Bengler, 2018; Sauer, Chavaillaz, & Wastell, 2016). Ideally, trust should be calibrated, meaning that drivers' trust should be aligned with the actual capabilities of the vehicle (Muir, 1994, Lee & See, 2004).

To understand how trust calibration can be achieved and the way in which trust impacts the use of automated vehicles, we need reliable trust metrics. Today, metrics are based primarily on questionnaires, in particular Jian, Bisantz, and Drury (2000) trust scale, which has been used in a broad range of studies (e.g., Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Hoff & Bashir, 2015; Gold et al., 2015). Nevertheless, self-reports do not provide continuous measurement and therefore cannot capture real-time changes in trust, like those associated with particular driving situations. Furthermore, questionnaires are hard to use in many applied settings, such as during real-world driving (Hergeth, Lorenz, Vilimek, & Krems, 2016). We here argue that analysis of drivers' gaze behaviour and skin conductance has the potential to provide objective, real-time measurement of trust.

Because drivers' reliance on automation depends to a large extent on trust (e.g., Carlson, Desai, Drury, Kwak, & Yanco, 2014), it is reasonable to assume that when trust is high, drivers will monitor non-driving-related tasks (NDRTs) more than when trust is low. This assumption is in line with findings in a driving simulator study by Beggiato and Krems (2013). These show that when participants had correct expectations regarding the behaviour of the car's automated functionalities, they trusted the vehicle and engaged more in a NDRT (Beggiato & Krems, 2013; Beggiato, 2015). This suggests that gaze behaviour has the potential to provide a reliable real-time measure of drivers' trust (Lee & See, 2004; Parasuraman & Manzey, 2010). However, the empirical evidence is mixed: While some studies find a relationship between trust and monitoring behaviour (e.g., Hergeth et al., 2016; Körber, Baseler, & Bengler, 2018; Walker, Verwey, & Martens, 2018), others do not (e.g., Gold et al., 2015).

Another way of measuring users' trust in automation is to analyse changes in skin conductance - the skin's ability to conduct an electrical current (Christie, 1981; Braithwaite, Watson, Jones, & Rowe, 2013). These changes, which are not under conscious control, are continuously modulated by the sympathetic nervous system, and are commonly referred to as Electrodermal Activity (EDA) (Critchley, 2002; Sequeira, Hot, Silvert, & Delplanque, 2009). Importantly, EDA is modulated by stress, most likely because it increases sweating (Costa, Roe, & Taillieu, 2001; Santarcangelo et al., 2012).

A number of recent studies have investigated the link between trust and EDA, showing that lower trust correlates with higher EDA. For example, Khawaji, Zhou, Chen, and Marcus (2015) investigated EDA as a possible indicator of interpersonal trust, showing that the EDA of users of a text-chat environment was strongly affected by trust and cognitive load. The link between trust and EDA was investigated further by Akash, Hu, Jain, and Reid (2018). In this study, participants were asked to evaluate the performance of a sensor, designed to detect on-road obstacles. After each of a series of trials, participants were asked if they trusted or distrusted the result from the sensor (i.e., presence or absence of an obstacle on the road). Participants' responses were followed by feedback, showing whether the response was correct or incorrect. The authors measured participants' EEG response signals and EDA during each trial, and built a trust model based on these values. They concluded that physiological measures were promising real-time indicators of participants' trust in automation (Akash et al., 2018). Morris and colleagues (2017) used a simulated autonomous vehicle to test participants' trust and EDA during risky and safe driving situations. The authors showed that, in the risky scenario, participants reported low trust towards the vehicle, and also showed an increase in EDA (Morris, Erno, & Pilcher, 2017). In summary, EDA seems a good indicator of emotional activation (Sequeira et al., 2009), revealing something of the emotions, actions, thoughts and perceptions of the participants in various studies (Cacioppo, Tassinary, & Berntson, 2007).

An earlier pilot investigation studied the potential of eye-tracking technology for the continuous measurement of trust. This showed that users' gaze behaviour provides important information on their trust towards an automated vehicle (Walker et al., 2018). The present study is a follow-up that explored whether combining gaze behaviour and EDA could provide, in real-time, a more reliable metric of drivers' trust. To that end, we simulated Fully Automated driving (for a taxonomy, see SAE, 2014) by projecting in a driving simulator forward looking videos recorded from a driving car. The decision to project videos in the simulator, instead of using a single monitor, was based on the rationale that the simulator grants a higher level of immersion. Research shows that the extent of the visual field, an objective aspect of "immersivity", positively influence participants' sense of presence – the feeling of being truly immersed in the virtual environment (Lessiter, Freeman, Keogh, & Davidoff, 2001; Slater, Usoh, & Steed, 1994) – and their driving behaviour (Risto & Martens, 2014). For these reasons, another recent study has used an experimental setup similar to our own (i.e., Boelhouwer, van den Beukel, van der Voort, & Martens, 2019).

In the present experiment, participants were divided into a "Perfect Vehicle" and a "Poor Vehicle" group. Participants were instructed to imagine being in a Fully Automated car and to perform a NDRT, but only if they trusted the car's behaviour. The experiment consisted of three driving phases. In Phase 1 and 2, the "automated" car performed differently for the two groups, handling the driving task perfectly for the Perfect Vehicle group, and poorly for the Poor Vehicle group. In Phase 3, both groups viewed videos of the car behaving poorly on the road. We expected to find a strong association between monitoring behaviour, EDA and self-reported trust: The more drivers trust the system, the less they view the road, the more attention they pay to the secondary task, and the lower their EDA. We also hypothesized that combining gaze behaviour and EDA would provide a better indication of drivers' trust, than either measure on its own.

2. Methods

2.1. Participants

Thirty-six participants, all students of the University of Twente, were recruited. These joined the study in exchange for study credits or money (6 euro). Ten participants were excluded from the analysis, due to poor eye-tracking or EDA data quality. The final sample consisted of 16 males and 10 females, all between 19 and 36 years of age (M = 24.27, SD = 4.37). None of the participants reported previous experience with automated vehicles, either as drivers or passengers. All participants had a driver's licence for at least two years, reported normal or corrected to normal (i.e., lenses) vision, did not wear glasses, and did not commonly suffer from motion sickness. The study was approved by the ethics board of the Faculty of Behavioural, Management and Social Sciences.

2.2. Video material

Fully Automated driving was simulated using two-minute videos in the University of Twente driving simulator. Videos were all filmed in the surroundings of the university's campus, through a Go-Pro Hero 4 Session camera, placed centrally on the hood of an Audi A4. The same material had been used in a previous study (i.e., Walker et al., 2018).

2.3. Tasks

Participants were assigned to one of two groups: As in Walker et al. (2018), the Perfect Vehicle group viewed forward looking videos of a car handling the driving task perfectly (i.e., it braked comfortably in front of crossing pedestrians or cyclists, and always kept in lane). The Poor Vehicle group viewed videos of a car struggling with the driving task, with a tendency to brake abruptly in front of crossing pedestrians or cyclists, and to drift towards the centre of the road. Perfect Vehicle and Poor Vehicle participants both encountered one crossing pedestrian or cyclist per video. Although group assignment was random, the randomization procedure ensured that the two groups had the same number of participants (i.e., sixteen per group), a similar gender make-up, and a similar distribution of initial trust levels.

During simulated automated driving, participants in both groups performed the visual Surrogate Reference Task (SuRT) (ISO 14198, 2012), a task commonly used to simulate NDRT during automated driving (e.g., Radlmayr, Gold, Lorenz, Farid, & Bengler, 2014; Hergeth et al., 2016) and in Adaptive Cruise Control (ACC) situations (e.g., Beggiato & Krems, 2013). Here, participants were to select, by tapping on a touch screen, the larger circle (diameter 47 pixels) among 49 distractor circles (diameter 40 pixels). Participants started a new trial by tapping the "start" button, displayed at the bottom of the screen at the end of each trial (see Fig. 1). The task was shown on an Apple iPad tablet with no time limit for completion. The placement of the tablet required participants to allocate their visual attention completely away from the road. Thus, participants could not use peripheral vision to watch the road while performing the task. The visual angle between the forward view on the road and the secondary task display amounted to about 12°.

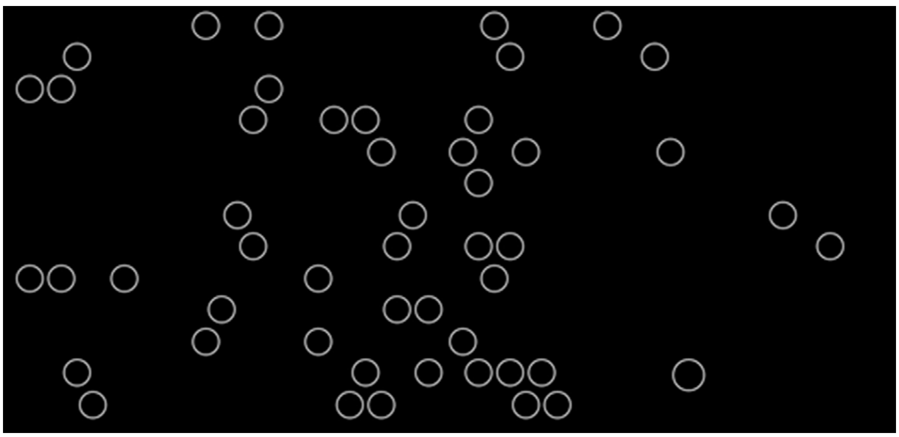

Fig. 1. The Surrogate Reference Task (SuRT) – trial example. Participants were instructed to search and tap on the larger circle, displayed among 49 distractor circles.

2.4. Measures

2.4.1. Self-reported trust

Participants' trust towards the simulated vehicle was measured through a modified version of the Empirically Determined (ED) Trust Scale (Jian et al., 2000). As in Verberne, Ham, and Midden (2012) and Walker et al. (2018), the modified version involved a 7-point Likert scale (1 = totally disagree; 7 = totally agree) indicating level of agreement with seven statements. The higher the score, the higher the trust. Trust scores were collected before the study, to ensure that participants in each group were equally distributed in terms of their initial trust, and also after every phase of the experiment. Cronbach's alpha of this scale was 0.90, 0.83, 0.87, 0.92 for the pre-test and after every phase, respectively.

2.4.2. Gaze behaviour

Participants' eye movements were recorded through a head-mounted mobile eye-tracker. Tobii Pro Glasses 2 is equipped with two cameras for each eye, and a wide-angle full HD scene camera to record the external world. The glasses were connected to a recording unit, and wirelessly to a Dell tablet, running the Windows 10 operating system and the Tobii Pro Glasses Controller software (Tobii Pro Glasses 2, 2018). Participants' gaze behaviour was extracted and analysed using Tobii Pro Lab software.

As shown in Fig. 2, we assessed participant fixations on two areas of interest (AOI): 1) the road, i.e., the central section of the driving simulator screen (size 2600 × 1950 mm); 2) the tablet (size 240 × 186 mm), on which the SuRT was presented. Fixations that fell outside the two AOIs were treated as "fixations on other area". Outcome variables were fixation frequency (i.e., percentage of fixations in each AOI), and fixation duration (i.e., percentage of time spent viewing each AOI).

2.4.3. Electrodermal Activity (EDA)

EDA, measured in microsiemens (μS), was recorded using the Empatica E4 wristband. The E4 is a wearable wireless watch-like device designed for continuous physiological data acquisition. EDA data was acquired through two ventral wrist electrodes, and was sampled at 4 Hz (Empatica E4 wristband, 2018). Given that participants had to use their right hand to perform the NDRT (see Fig. 2), the E4 wristband was placed on participants' left wrist. EDA values were z-transformed to compensate for higher inter-individual variability of EDA (Ben-Shakhar, 1985; 1987).

2.5. Driving simulator

The University of Twente driving simulator consists of a mock-up equipped with steering wheel, pedals and indicators. Psychopy software (Peirce, 2009) was used to display the videos on a screen of 7800 × 1950 mm. The screen has a total resolution of 3072 * 768 pixels (~10ppi).

2.6. Procedure

Participants completed the pre-test questionnaire two or three days before the experiment. It consisted of demographic questions and the trust scale, used to measure participants' initial trust in automated cars. On the testing day, participants

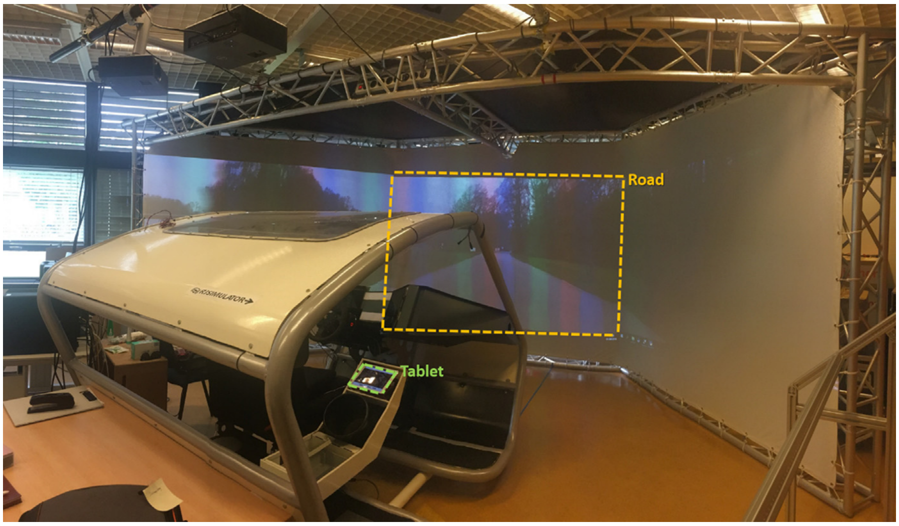

Fig. 2. Areas of Interest (AOI). We assessed participants' fixations on the central screen (yellow) and the tablet (green), where the NDRT was presented. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

were welcomed and told that they would view videos while sitting in the simulator mock-up. After filling in an informed consent form, participants put on the E4 wristband and the eye-tracker. The latter was then calibrated following Tobii's recommendations (Tobii Pro Glasses 2 – Calibrating the participant, 2018). After calibration, participants sat down in the driving simulator. On screen instructions informed them that the experiment would be divided into 3 phases, and instructed them to imagine being in a Fully Automated car at all times. Participants were told to perform the NDRT, but only if they trusted the behaviour of the car. Participants practiced the NDRT before starting Phase 1. Each phase was composed of two videos. The order of the videos was counterbalanced across participants.

Phase 1 served as a habituation period, in which participants could develop their own expectations towards the self-driving car: High trust for the Perfect Vehicle group, since the car behaved perfectly; low trust for the Poor Vehicle group, since the car behaved poorly. In this phase participants could also get used to the simulator and the NDRT. After this, and each further phase, participants filled out the trust questionnaire (see 2.4.1). Importantly, although the items of the questionnaire were the same as in the pre-test, they were rephrased to refer to participants' trust in the car they had just experienced, and not their general trust towards automated vehicles. In Phase 2, the car behaved in the same way as in Phase 1, though the driving scenarios were different. In Phase 3, both groups viewed videos in which the car handled the driving task poorly (the same videos for each group). After filling out the last trust questionnaire, participants left the mock-up, and were helped to remove the wrist band and the eye-tracker. Finally, they filled in an exit form, through which they could point out suggestions for the improvement of the study. The entire experiment lasted approximately 45 minutes.

3. Results

3.1. Self-reported trust

An independent samples *t*-test showed that the pre-test trust level was not different in the Perfect Vehicle group (M = 4.02, SD = 1.29) and the Poor Vehicle group (M = 4.09, SD = 1.31), $t(24) = -0.129$, $p = .898$. Self-reported trust after experience with the simulator was analysed using a 4 × 2 mixed Analysis of Variance (ANOVA), with Phase as the within-subjects factor (Pre/Phase 1/Phase 2/Phase 3) and Group as the between-subjects factor (Perfect Vehicle group/Poor Vehicle group). Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(2) = 8.574$, p = .128. Results showed a significant main effect of Phase (F(3, 72) = 6.21, $p = .001$, $\eta_p^2 = 0.21$), and an interaction between Phase and Group (F(3,72) = 20.00, $p < .001$, $\eta_p^2 = 0.46$, Fig. 3).

The significant interaction effect between Phase and Group was further analysed using independent samples two-tailed t-tests for each phase. As expected, the Perfect Vehicle group (M = 4.88, SD = 0.60) showed higher trust levels after Phase 1 than the Poor Vehicle group (M = 3.35, SD = 0.98), $t(24) = 4.797$, $p < .001$, $d = 1.88$. Similarly, at the end of Phase 2, the Perfect Vehicle group (M = 5.10, SD = 0.57) again showed higher trust levels than the Poor Vehicle group (M = 3.48, SD = 1.11), $t(24) = 4.672$, $p < .001$, $d = 1.84$. At the end of Phase 3, by contrast, the Perfect Vehicle group (M = 2.87, SD = 1.16) showed significantly lower trust levels than the Poor Vehicle group (M = 4.01, SD = 1.26), $t(24) = -2.40$, $p = .025$, $d = -0.94$.

Paired samples two-tailed t-tests were used to investigate the effect of phase within each group. Compared to their initial trust (i.e., "Pre" in Fig. 3), the Perfect Vehicle group reported higher trust after Phase 1 ($t(12) = -3.449$, $p = .005$, $d = -0.96$) and Phase 2 ($t(12) = -3.651$, $p = .003$, $d = -1.01$) and lower trust after Phase 3 ($t(12) = 3.681$, $p = .003$, $d = 1.02$). Trust increased significantly from Phase 1 to Phase 2 ($t(12) = -2.343$, $p = .037$, $d = -0.65$) and dropped after Phase 3 ($t(12) = 8.713$, $p < .001$, $d = 2.42$). In the Poor Vehicle group, there were no significant differences between initial trust levels and the levels reported after Phase 1 ($t(12) = 1.680$, $p = .119$), Phase 2 ($t(12) = 1.723$, $p = .110$), and Phase 3 ($t(12) = 0.218$, $p = .831$). Surprisingly, however, trust was significantly higher in Phase 3 than in Phase 2 ($t(12) = -2.281$, $p = .042$, $d = -0.63$) (see Fig. 3).

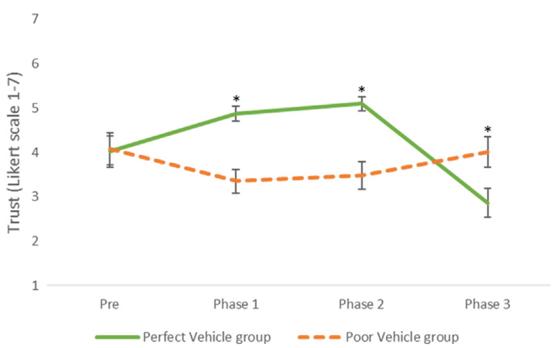

Fig. 3. Mean self-reported trust scores per phase and group. The Perfect Vehicle group reported higher trust scores during Phase 1 and 2. The Poor Vehicle group reported higher trust during Phase 3. Interestingly, trust fell from Phase 2 to Phase 3 in the Perfect Vehicle but increased in the Poor Vehicle group. Error bars represent standard error of means, and each * indicates a significant difference between the two groups ($p <= .05$).

3.1.1. Gaze behaviour

Mann-Whitney U tests showed that there was no significant difference in fixation frequency and duration between the Perfect Vehicle and the Poor Vehicle groups in Phase 1 (Frequency – to the Road: U = 66, p = .362; to the NDRT: U = 60, p = .223; Duration – to the Road: U = 62, p = .264; to the NDRT: U = 54, p = .125, see Fig. 4). As expected, in Phase 2 fixation frequency and duration differed significantly between the groups, with participants in the Perfect Vehicle group making fewer fixations (U = 46, p = .05) and spending less time (U = 42, p = .029) viewing the road. Fixation frequency (U = 38, p = .016) and time (U = 35.5, p = .01) spent on the NDRT were in line with these results. As hypothesized, no significant differences were found between the two groups during Phase 3. (Frequency - Road: U = 61, p = .243; NDRT: U = 68.5, p = .418; Duration - Road: U = 61, p = .243; NDRT: U = 66.5, p = .362). Means and standard deviations are reported in Tables 1 and 2.

Within-group differences between Phase 1, Phase 2 and Phase 3 were assessed using a Friedman test. In the Perfect Vehicle group, gaze frequency (to the Road: χ^2 = 12.275, p = .002; to the NDRT: χ^2 = 10.308, p = .006) differed between the three phases. Wilcoxon Signed Ranks tests (see Table 1) showed that the Perfect Vehicle group made significantly fewer fixations on the road in Phase 2 than in Phase 1 (p = .005). In Phase 3, in contrast, the group fixated more frequently on the road than in

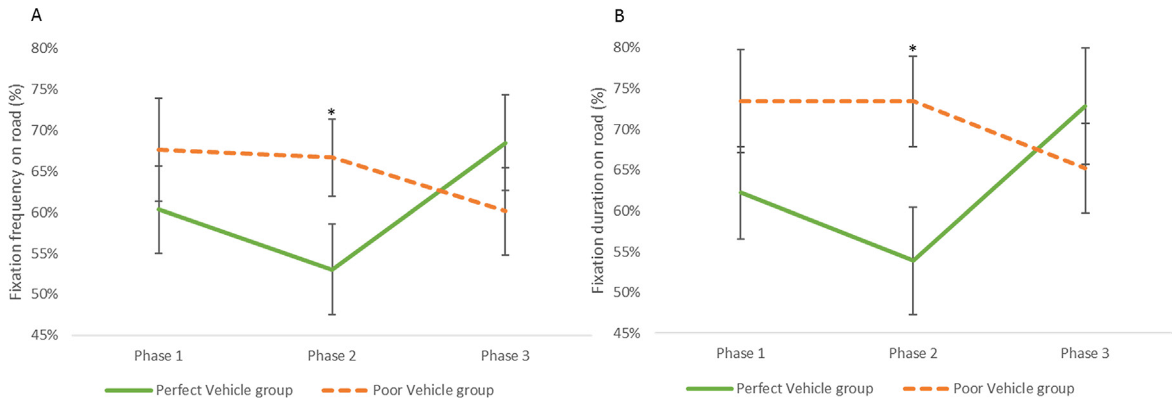

Fig. 4. Participants monitoring behaviour. (A) Fixation frequency on road; (B) Fixation duration on road. During Phase 2, participants in the Perfect Vehicle group made more fixations and spent more time viewing the road. In Phase 1 and 3, there was no difference between the two groups. Error bars represent standard error of means, and each * indicates a significant difference between the two groups (p <= .05).

Table 1
Fixation frequency. Wilcoxon Signed Ranks test, in this and following tables, was used for pairwise comparisons.

		Phase 1 (M, SD)	Phase 2 (M, SD)	Phase 3 (M, SD)	Phase 1 VS Phase 2	Phase 1 VS Phase 3	Phase 2 VS Phase 3
Perfect Vehicle group	Road	M = 60%, SD = 0.19	M = 53%, SD = 0.20	M = 68%, SD = 0.21	Z = −2.825, p = .005*	Z = −2.201, p = .028*	Z = −3.110, p = .002*
	NDRT	M = 28%, SD = 0.14	M = 30%, SD = 0.16	M = 16%, SD = 0.19	Z = −1.013, p = .311	Z = −2.551, p = .011*	Z = −2.900, p = .004*
Poor Vehicle group	Road	M = 68%, SD = 0.23	M = 67%, SD = 0.17	M = 60%, SD = 0.19	Z = −0.524, p = .600	Z = −1.992, p = .046*	Z = −1.726, p = .084
	NDRT	M = 21%, SD = 0.19	M = 14%, SD = 0.15	M = 22%, SD = 0.18	Z = −1.452, p = .147	Z = −0.510, p = .610	Z = −1.961, p = .050*

* Indicates statistical significance (p <=.05).

Table 2
Fixation duration.

		Phase 1 (M, SD)	Phase 2 (M, SD)	Phase 3 (M, SD)	Phase 1 VS Phase 2	Phase 1 VS Phase 3	Phase 2 VS Phase 3
Perfect Vehicle group	Road	M = 62%, SD = 0.20	M = 54%, SD = 0.24	M = 73%, SD = 0.26	Z = −2.589, p = .010*	Z = −2.201, p = .028*	Z = −3.180, p = .001*
	NDRT	M = 30%, SD = 0.17	M = 35%, SD = 0.21	M = 19%, SD = 0.23	Z = −1.642, p = .101	Z = −2.551, p = .011*	Z = −3.180, p = .001*
Poor Vehicle group	Road	M = 73%, SD = 0.23	M = 73%, SD = 0.20	M = 65%, SD = 0.20	Z = −0.245, p = .807	Z = −1.712, p = .087	Z = −1.608, p = .108
	NDRT	M = 19%, SD = 0.19	M = 14%, SD = 0.15	M = 23%, SD = 0.17	Z = −1.255, p = .209	Z = −1.020, p = .308	Z = −2.118, p = .034*

* Indicates statistical significance (p <= .05).

Phase 1 ($p = .028$) or Phase 2 ($p = .002$). Fixations on the NDRT followed an almost inverse pattern with no difference between Phase 1 and Phase 2, but significant differences between these two phases and Phase 3. In the latter, participants fixated less frequently on the NDRT (Phase 1, $p = .011$; Phase 2, $p = .004$). In the Poor Vehicle group, we observed no significant difference between Phase 1 and Phase 2 in terms of fixations on the road and fixations on the NDRT. In Phase 3, surprisingly, participants fixated less on the road than in Phase 1 ($p = .046$). This was not the case when comparing Phase 3 to Phase 2 ($p = .084$). Though, in Phase 3, participants of the Poor Vehicle group did fixate more on the NDRT than in Phase 2 ($p = .05$).

A Friedman test showed strong differences in fixation duration within each group (Road: $\chi^2 = 15.176$, $p = .001$; NDRT: $\chi^2 = 15.846$, $p < .001$), a pattern similar to the one observed with fixation frequencies. Wilcoxon Signed Ranks tests (see Table 2) showed that the Perfect Vehicle group spent less time viewing the road in Phase 2 than in Phase 1 ($p = .01$) and more time in Phase 3, than in either Phase 1 ($p = .028$) or Phase 2 ($p = .002$). No difference was found when comparing the time participants of the Perfect Vehicle group spent viewing the NDRT during Phase 1 and Phase 2. Nevertheless, the same group spent less time viewing the NDRT during Phase 3 than in either Phase 1 ($p = .011$) or Phase 2 ($p = .001$). In the Poor Vehicle group, participants spent more time viewing the NDRT in Phase 3 than in Phase 2 ($p = .034$). Otherwise, the Poor Vehicle group showed no significant differences between phases.

3.2. Electrodermal Activity (EDA)

As already reported for gaze behaviour, Mann-Whitney U tests revealed significant differences in EDA between the Perfect and Poor Vehicle groups. In Phase 1, although EDA may appear higher in the Perfect Vehicle group than in the Poor Vehicle group, this difference only approached significance ($U = 47$, $p = .057$, see Fig. 5). However, in Phase 2, EDA was significantly lower in the Perfect Vehicle group than in the Poor Vehicle group ($U = 42$, $p = .029$). As expected, there were no significant differences between the two groups in Phase 3, ($U = 77$, $p = .724$) (see Table 3).

In the Perfect Vehicle group there were no significant within-group differences between Phase 1, Phase 2 and Phase 3 ($\chi^2 = 1.846$, $p = .397$). In the Poor Vehicle group differences only approached significance ($\chi^2 = 5.692$, $p = .058$). Wilcoxon Signed Ranks tests (see Table 3) showed that members of the Poor Vehicle group had significantly lower EDA in Phase 1, than in either Phase 2 or Phase 3. No difference was found between Phase 2 and Phase 3 (see Table 3 and Fig. 5).

3.3. Correlations

Pearson correlations were used to further analyse the relationship between self-reported trust, monitoring behaviour and EDA. In all phases, trust scores collected at the end of each phase and pooled across the two groups correlated strongly with gaze behaviour. Combining the results of Phase 1 through 3 yielded similar results (see Table 4). In brief, the higher the trust, the less participants fixated the road, and the more they fixated on the screen displaying the NDRT (see Fig. 6).

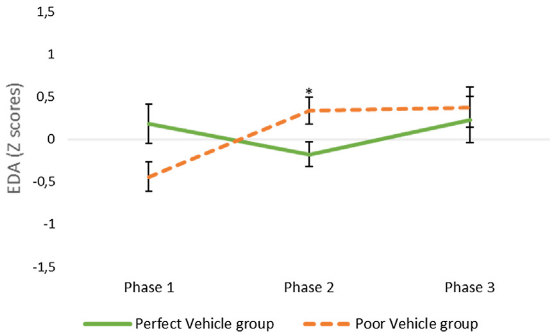

Fig. 5. Electrodermal Activity (EDA). During Phase 2 members of the Poor Vehicle group showed higher EDA than participants of the Perfect Vehicle group. In Phase 1 the difference between the two groups only approached significance. In Phase 3 there were no significant differences between the groups. Error bars represent standard error of means, and each * indicates a significant difference between the two groups ($p <= .05$).

Table 3
Electrodermal Activity (EDA).

	Phase 1 (M, SD)	Phase 2 (M, SD)	Phase 3 (M, SD)	Phase 1 VS Phase 2	Phase 1 VS Phase 3	Phase 2 VS Phase 3
Perfect Vehicle group	M = 0.18, SD = 0.83	M = −0.18, SD = 0.51	M = 0.23, SD = 0.98	Z = −0.943, $p = .345$	Z = −0.314, $p = .753$	Z = −1.433, $p = .152$
Poor Vehicle group	M = −0.44, SD = 0.63	M = 0.34, SD = 0.56	M = 0.37, SD = 0.84	Z = −2.062, $p = .039$*	Z = −1.992, $p = .046$*	Z = −0.314, $p = .753$

* Indicates statistical significance ($p <= .05$).

Table 4
Correlations between self-reported trust and gaze behaviour.

		Phase 1	Phase 2	Phase 3	Phases 1–3
Fixation frequency	Road	r = −0.532, p = .005*	r = −0.514, p = .007*	r = −0.578, p = .002*	r = −0.497, p = .010*
	NDRT	r = 0.606, p = .001*	r = 0.587, p = .001*	r = 0.555, p = .003*	r = 0.556, p = .003*
Fixation duration	Road	r = −0.599, p = .001*	r = −0.602, p = .001*	r = −0.625, p = .001*	r = −0.585, p = .002*
	NDRT	r = 0.651, p < .001*	r = 0.642, p < .001*	r = 0.592, p = .005*	r = 0.619, p = .001*

* Indicates statistical significance (p <= .05).

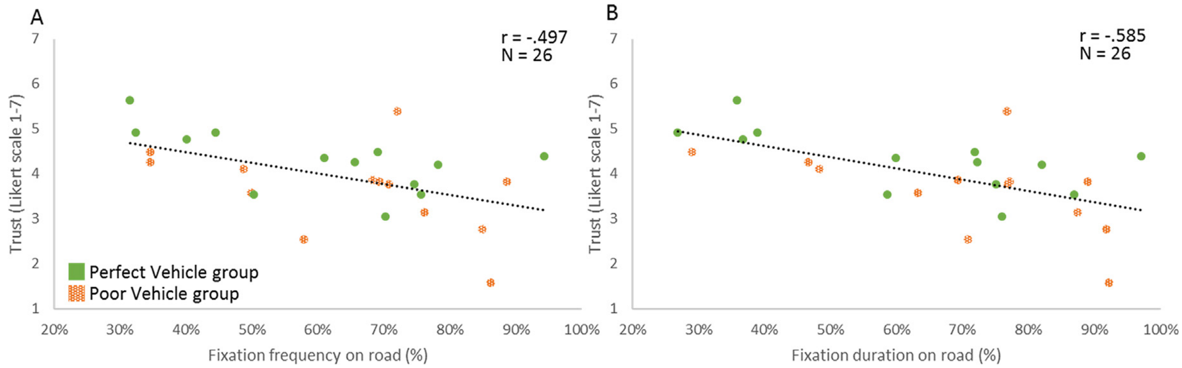

Fig. 6. Correlation between self-reported trust (1: low; 7: high) and monitoring behaviour; combined results of Phase 1, 2, and 3. Both graphs refer to on-road fixations: (A) fixation frequency; (B) Fixation duration. The higher the trust, the less participants monitored the road.

We found no correlation of EDA with trust scores, either in Phase 1 (r = 0.246, p = .225) or in Phase 3 (r = 0.028, p = .893). In Phase 2, however, there was a strong negative correlation (r = −0.575, p = .002) indicating that higher trust correlated with lower EDA (see Fig. 7).

Finally, we tested whether there were any correlations between EDA, fixation frequency and fixation duration. The only significant correlations were again in Phase 2, where participants' EDA correlated with fixation duration (Road: r = 0.443, p = .024; NDRT: r = −0.447, p = .022). The more time participants spent viewing the road, the higher their EDA (see Fig. 8).

3.4. Joint use of gaze duration and EDA

Following these results, we investigated whether combining Gaze Duration and EDA could provide a more reliable, real-time indicator of trust, than either measure on its own. To test this hypothesis, we compared the linear regression of

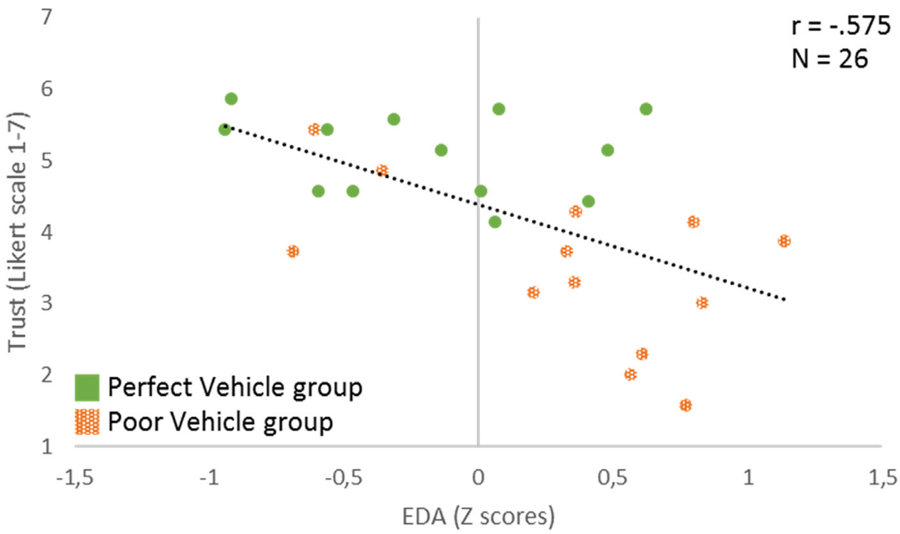

Fig. 7. Correlation between self-reported trust (1: low; 7: high) after Phase 2 and EDA observed during Phase 2. The higher the trust, the lower participants' EDA.

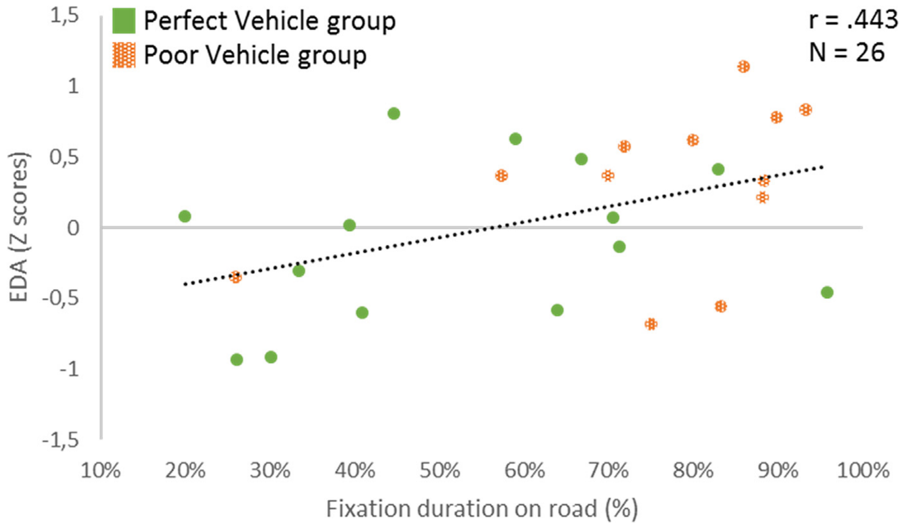

Fig. 8. Correlation between the individual participants' EDA values and fixation duration during Phase 2. The more time participants' spent viewing the road, the higher their EDA.

Table 5
Linear regression, Phase 2. B represents the unstandardized coefficient.

	B	R^2	*p*	F
On-Road fixation duration + EDA	Constant: 5.744 Duration: −2.184 EDA: −0.777	0.481	.001*	10.662
On-Road Fixation duration	Constant: 6.226 Duration: −3.040	0.363	.001*	13.675
EDA	Constant: 4.385 EDA: −1.166	0.330	.002*	11.843

* Indicates statistical significance ($p <= .05$).

self-reported trust in Phase 2 with gaze duration and EDA, against regressions using these variables separately. The results confirm that the combination of the two indicators predicts self-reported trust better than either individually (see Table 5).

4. Discussion

In this study we investigated the potential of gaze behaviour and Electrodermal Activity (EDA), singly and combined, as real-time measures of trust in automation. As expected, percentage of on-road fixation frequency and duration were negatively related to self-reported trust, while gaze behaviour on the NDRT was positively correlated to self-reported trust: The higher the drivers' self-reported trust, the less time they spent viewing the road (see Table 4 and Fig. 6). These results confirm the negative relationship between trust and drivers' on-road monitoring behaviour during automated driving, already observed in previous studies (i.e., Hergeth et al., 2016; Körber et al., 2018; Walker et al., 2018).

The study was organized into three phases. In Phases 1 and 2, the Perfect Vehicle group viewed videos of a simulated automated car behaving well on the road. Conversely, during the same phases, the Poor Vehicle group viewed videos of the car struggling with the driving task. In Phase 3, both groups viewed the same videos of the car behaving poorly on the road. In Phases 1 and 2 we expected participants' gaze behaviour, EDA and self-reported trust to differ between the Perfect and the Poor Vehicle group. In Phase 3 we expected to find no differences between the two groups.

Gaze behaviour differed between the two groups in Phase 2, but not in Phases 1 and 3, where the Perfect and Poor Vehicle groups displayed the same behaviour (see Fig. 4). Trust develops over time (Lee & See, 2004). Once the two groups had experienced the behaviour of their respective vehicles, trust developed differently within each group (see Fig. 3), and their gaze behaviour changed accordingly.

The results for EDA were similar to those for gaze behaviour. In Phase 1 there were no significant inter-group differences. In Phase 2, the Perfect Vehicle group showed significantly lower EDA than the Poor Vehicle group. No difference was found in Phase 3 where, as expected, EDA was high in both groups (see Table 3 and Fig. 5). These results confirm the findings of Khawaji et al. (2015) and Morris et al. (2017): Demanding and stressful events increase individuals' EDA. In Phase 2, EDA correlated strongly with self-reported trust: The higher the trust, the lower participants' EDA (see Fig. 7). Along the same lines, the more time drivers spent viewing the road during Phase 2, the higher their EDA (see Fig. 8). Unfortunately, in Phase 1 and 3, no correlations were found between EDA and self-reported trust. Previous studies have shown that too often, data collected through wearable and stationary sensors do not match (e.g., Ollander, Godin, Campagne, & Chararbonnier, 2016). It is possible, therefore, that our choice of instrument for the collection of the EDA data - the E4 wrist band – may have influenced our results. Future research should investigate whether more precise physiological measurements would lead to stronger effects.

Our results show that while gaze behaviour and EDA both provide useful real-time information about trust, they are both noisy measurements. Importantly, our study provides preliminary indications that combining the two measures may provide a more precise indicator of user trust than either measure individually (see Table 5). Given, however, that this result is derived from a single phase (i.e., Phase 2) of a single study, it requires validation in future work.

Our results, if confirmed, suggest possible applications in real-life driving scenarios. In its CT6 model (Cruise, 2018) Cadillac uses head-tracking technology to track drivers' head movements and infer their alertness, providing appropriate audio and visual feedback when the driver is not paying attention to the road. In-vehicle eye-tracking technology could be used in a similar way; not only to monitor drivers' physiological state, but also to tailor messages and alerts, in real-time, to their trust towards the automated system.

Our study has a number of limitations. Due to the video-based nature of the experiment, participants knew they had no way of disengaging the automated system and take back control of the vehicle manually. In a real-world scenario, it is unlikely that drivers would rely on the automated vehicle if they thought it was not behaving safely. Importantly, though, we found our results in a situation that is not even actually threatening, that is, in a driving simulator. In real-world automated driving the effects may be larger. Thus, although the "absence of risk" may have affected study results, the data still suggest that our methodology can be validly used for the investigation of users' trust. Furthermore, it is important to note that, in Phase 1, we were expecting gaze behaviour and EDA of the two groups to differ. This was not the case. Phase 1 served as a habituation period. It is thus possible that participants used a big part of Phase 1 to get used to the task, and this may explain the absence of differences between the two groups. Moreover, while drivers' gaze behaviour could be easily tracked in an actual vehicle through fixed cameras, measuring continuously EDA, without annoying the vehicle user, may be a harder challenge. Finally, since analysing task performance was outside the scope of this study, we did not verify whether our participants were left or right handed. Although we doubt that participants' handedness affected any of our results, this possibility cannot be ruled out.

Despite these limitations, analysis of participants' self-reported trust confirmed the success of our manipulation and suggested new ideas for future studies. After Phase 1 and 2, the Perfect Vehicle group reported higher trust than the Poor Vehicle group. These results are in line with previous studies, showing that users' initial trust in automation can be modified by training and experience (Bahner, Hüper, & Manzey, 2008; Sauer et al., 2016; Walker et al., 2018). Nevertheless, an interesting result was observed in Phase 3, where both groups viewed videos of the car struggling with the driving task. Although both groups viewed the same videos in this phase, the Perfect Vehicle group's trust dropped while in the Poor Vehicle group it increased, actually reaching a higher level than in the Perfect Vehicle group. These findings show not just that unexpected system failures can highly disrupt trust in automation, but also that humans can deal with bad technology, so long as it is reasonably predictable. This result is in line with the literature (e.g., Lee & See, 2004; Rempel et al., 1985; Beggiato & Krems, 2013; Beggiato, 2015). Beggiato and Krems (2013) show, among others, that well-known system failures do not negatively affect trust in automation.

5. Conclusions

Gaze behaviour and EDA appear to be promising indicators of drivers' trust in automation. Although both measurements can be used independently, our results suggest that the combination of the two provides a better indication of drivers' trust in automation than either individually. Overall, our study represents one more step towards the development of reliable, continuous and objective measurements of drivers' trust in automated vehicles. Such measurements are particularly important for the assessment of changes in real-time trust, especially in real-world scenarios, where measures, such as questionnaires, are harder to deploy. This would, for example, make it possible to distinguish trust levels associated with specific driving situations.

Importantly, although the metrics presented in this article (e.g., percentage of road fixations) were calculated post-hoc, it is of course possible to develop a real-time algorithm based on the same metrics, calculated over a moving time window. If in a particular time frame, percentage of road fixations and EDA would be below or above a specific threshold (calculated based on system performance), the algorithm would assume that user's trust is not calibrated. The results could be used to provide drivers with visual and audio feedback, designed to improve the alignment between driver trust and the real capabilities of the vehicle.

Future work is still needed for the development of reliable real-time measures of drivers' trust in automated vehicles. The link between trust, gaze behaviour and EDA needs to be investigated further, together with other objective metrics, such as heart rate. Our study represents a step forward in the development of such measures, and thus in the understanding of how trust in automation affects human behaviour.

Funding

This research is supported by the Dutch Domain Applied and Engineering Sciences, which is part of the Netherlands Organisation for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs (project number 14896).

Data availability statement

The data collected during the study are available at https://doi.org/10.17605/OSF.IO/8JE5C.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

References

Akash, K., Hu, W. L., Jain, N., & Reid, T. (2018). A Classification Model for Sensing Human Trust in Machines Using EEG and GSR. arXiv preprint arXiv:1803.09861. Retrieved from: https://arxiv.org/abs/1803.09861.

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

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

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. (2015). Changes in motivational and higher-level cognitive processes when interacting with in-vehicle automation (doctoral dissertation) Retrieved from:. *University of Technology Chemnitz* http://monarch.qucosa.de/api/qucosa%3A20246/attachment/ATT-0/.

Ben-Shakhar, G. (1985). Standardization within individuals: A simple method to neutralize individual differences in skin conductance. *Psychophysiology, 22*(3), 292–299. https://doi.org/10.1111/j.1469-8986.1985.tb01603.x.

Ben-Shakhar, G. (1987). The correction of psychophysiological measures for individual differences in responsivity should be based on typical response parameters: A reply to Stemmler. *Psychophysiology, 24*(2), 247–249. https://doi.org/10.1111/j.1469-8986.1987.tb00287.x.

Boelhouwer, A., van den Beukel, A. P., van der Voort, M. C., & Martens, M. H. (2019). Should I take over? Does system knowledge help drivers in making take-over decisions while driving a partially automated car? *Transportation Research Part F: Traffic Psychology and Behaviour, 60*, 669–684. https://doi.org/10.1016/j.trf.2018.11.016.

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments Retrieved from:. *Psychophysiology, 49*(1), 1017–1034.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). (2007). *Handbook of psychophysiology*. https://doi.org/10.1017/S0033291707001201.

Cadillac Super Cruise (2018). Retrieved from https://www.cadillac.com/content/dam/cadillac/na/us/english/index/ownership/technology/supercruise/pdfs/2018-cad-ct6-supercruise-personalization.pdf.

Carlson, M. S., Desai, M., Drury, J. L., Kwak, H., & Yanco, H. A. (2014). Identifying factors that influence trust in automated cars and medical diagnosis systems Retrieved from:. In *AAAI Symposium on The Intersection of Robust Intelligence and Trust in Autonomous Systems* (pp. 20–27). Palo Alto, CA: Association for the Advancement of Artificial Intelligence.

Choi, J. K., & Ji, Y. G. (2015). Investigating the Importance of Trust on Adopting an Autonomous Vehicle. *International Journal of Human-Computer Interaction, 31*(10), 692–702. https://doi.org/10.1080/10447318.2015.1070549.

Christie, M. J. (1981). Electrodermal activity in the 1980s: A review. *Journal of the Royal Society of Medicine, 74*(8), 616–622. https://doi.org/10.1177/014107688107400812.

Costa, A. C., Roe, R. A., & Taillieu, T. (2001). Trust within teams: The relation with performance effectiveness. *European journal of work and organizational psychology, 10*(3), 225–244. https://doi.org/10.1080/13594320143000654.

Critchley, H. D. (2002). Electrodermal responses: What happens in the brain. *The Neuroscientist, 8*(2), 132–142. https://doi.org/10.1177/107385840200800209.

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

Empatica E4 wristband (2018, September 24). Retrieved from: https://support.empatica.com/hc/en-us/articles/202581999-E4-wristband-technical-specifications.

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.

Guenzi, P., & Georges, L. (2010). Interpersonal trust in commercial relationships: Antecedents and consequences of customer trust in the salesperson. *European Journal of Marketing, 44*(1/2), 114–138. https://doi.org/10.1108/03090561011008637.

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

ISO 14198. (2012). PD ISO/TS 14198:2012-Road vehicles - Ergonomic aspects of transport information and control systems - Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems: BSI. Retrieved from: https://www.iso.org/standard/54496.html.

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

Khawaji, A., Zhou, J., Chen, F., & Marcus, N. (2015). Using galvanic skin response (GSR) to measure trust and cognitive load in the text-chat environment. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1989–1994). ACM. https://doi.org/10.1145/2702613.2732766.

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

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

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

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators and virtual environments, 10*(3), 282–297. https://doi.org/10.1162/105474601300343612.

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

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

Ollander, S., Godin, C., Campagne, A., & Charbonnier, S. (2016). A comparison of wearable and stationary sensors for stress detection. *Proceedings IEEE International Conference on Systems, Man, and Cybernetics, October*.. https://doi.org/10.1109/SMC.2016.7844917.

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

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

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*, 10. https://doi.org/10.3389/neuro.11.010.2008.

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

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

Risto, M., & Martens, M. H. (2014). Driver headway choice: A comparison between driving simulator and real-road driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 25*, 1–9. https://doi.org/10.1016/j.trf.2014.05.001.

Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist, 35*(1), 1. https://doi.org/10.1037/0003-066X.35.1.1.

Santarcangelo, E. L., Paoletti, G., Balocchi, R., Scattina, E., Ghelarducci, B., & Varanini, M. (2012). Watching neutral and threatening movies: Subjective experience and autonomic responses in subjects with different hypnotizability levels. *International Journal of Psychophysiology, 84*(1), 59–64. https://doi.org/10.1016/j.ijpsycho.2012.01.010.

Sauer, J., Chavaillaz, A., & Wastell, D. (2016). 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.

Sequeira, H., Hot, P., Silvert, L., & Delplanque, S. (2009). Electrical autonomic correlates of emotion. *International Journal of Psychophysiology, 71*(1), 50–56. https://doi.org/10.1016/j.ijpsycho.2008.07.009.

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments, 3*(2), 130–144. https://doi.org/10.1162/pres.1994.3.2.130.

Taxonomy - S. A. E. Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. Technical report, SAE International, 2014. https://doi.org/10.4271/J3016_201401.

Tobii Pro Glasses 2 (2018, September 24). Retrieved from: https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/.

Tobii Pro Glasses 2 - Calibrating the participant (2018, September 24). Retrieved from: https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-glasses-2-user-manual.pdf/?v=1.83.

Urmson, C., & Whittaker, W. (2008). Self-driving cars and the urban challenge. *IEEE Intelligent Systems, 23*(2). https://doi.org/10.1109/MIS.2008.34.

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

Walker, F., Boelhouwer, A., Alkim, T., Verwey, W., & Martens, M. (2018). Changes in trust after driving Level 2 automated cars. *Journal of Advanced Transportation*. https://doi.org/10.1155/2018/1045186.

Walker, F., Verwey, W., & Martens, M. (2018). Gaze behaviour as a measure of trust automated vehicles. *Proceedings of the 6th Humanist Conference, the Hague*.