The influence of experienced severe road traffic accidents on take-over reactions and non-driving-related tasks in an automated driving simulator study

Article *in* Accident; Analysis and Prevention · November 2021

DOI: 10.1016/j.aap.2021.106408

Accident Analysis and Prevention 162 (2021) 106408

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

The influence of experienced severe road traffic accidents on take-over reactions and non-driving-related tasks in an automated driving simulator study

Klemens Weigl[a,b,*], Clemens Schartmüller[a,c], Philipp Wintersberger[d], Marco Steinhauser[b], Andreas Riener[a]

[a] Human-Computer Interaction Group, Technische Hochschule Ingolstadt, Germany
[b] Department of Psychology, Catholic University of Eichstätt-Ingolstadt, Germany
[c] Johannes Kepler University Linz, Austria
[d] Institute of Visual Computing and Human-Centered Technology, Technische Universität Wien, Austria

ARTICLE INFO

Keywords:
Severe RTA experiences
Automated driving SAE Level 3
TOR
Non-driving-related tasks
PTSD
Depression and psychological resilience

ABSTRACT

Road traffic accidents (RTAs) are an ever-existing threat to all road users. Automated vehicles (AVs; SAE Level 3–5) are developed in many countries. They are promoted with numerous benefits such as increased safety yielding less RTAs, less congestion, less greenhouse gas emissions, and the possibility of enabling non-driving related tasks (NDRTs). However, there has been no study which has investigated different NDRT conditions, while comparing participants who experienced a severe RTA in the past with those who experienced no RTA. Therefore, we conducted a driving simulator study (N = 53) and compared two NDRT conditions (i.e., auditory-speech (ASD) vs. heads-up display (HUD)) and an accident (26 participants) with a non-accident group (27; between-subjects design). Although our results did not reveal any interaction effect, and no group difference between the accident and the non-accident group on NDRT, take-over request (TOR), and driving performance, we uncovered for both groups better performances for the HUD condition, whereas a lower cognitive workload was reported for the ASD condition. Nevertheless, there was no difference for technology trust between the two conditions. Albeit we observed higher self-ratings of PTSD symptoms for the accident than for the non-accident group, there were no group differences on depression and psychological resilience self-ratings. We conclude that severe RTA experiences do not undermine NDRT, TOR, and driving performance in a SAE Level 3 driving simulator study, although PTSD symptoms after an RTA may affect the psychological wellbeing.

1. Introduction

Road traffic accidents (RTAs) pose a major global challenge towards more road safety on the way to Vision Zero (i.e., zero fatalities and zero serious injuries within the road transport system; Belin et al., 1997; Johansson, 2009). In higher developed countries, RTAs have been decreasing in the last decades (e.g., due to improvements in medical care, technology, vehicle and road design, better driver education and training; Naqvi et al., 2020). Nevertheless, approximately 1.35 million people die per year because of road crashes, which are amongst the top 10 leading causes of deaths in lower, lower-middle, and upper-middle income countries, and the leading cause of death worldwide for children and young adults aged 5–29 years (Bener, 2005; Manyara, 2016; Museru et al., 2002; Singh, 2017; WHO 2020a; b). RTAs are caused by *human factors* such as over speeding, distraction (e.g., by phone use), not wearing a seat belt, reckless and drunken driving, drug abuse, and disobeying traffic rules, by *nature's and external influences* (i.e., animal crossings, extreme weather conditions, poor roads, etc.), or because of *technological issues*, for example, tire blowouts (Ansari et al., 2000; Gopalakrishnan, 2012; Mayou et al., 1993; Valent et al., 2002; Vorko-Jović et al., 2006). For the people involved, RTAs can have substantial psychological consequences. Mayou et al. (1993) investigated the psychological impact of severe RTAs and observed the *acute stress syndrome* in one fifth of the subjects (i.e., mood disturbance and horrific memories of the accident), whereas increased anxiety, depression, and post-traumatic symptoms over 12 months after the accident occurred even

* Corresponding author at: Technische Hochschule Ingolstadt (THI), Paradeplatz 13, 85049 Ingolstadt, Germany.
E-mail address: klemens.weigl@gmail.com (K. Weigl).

https://doi.org/10.1016/j.aap.2021.106408
Received 2 April 2021; Received in revised form 26 August 2021; Accepted 20 September 2021

more often. They also found that one tenth of the patients suffered from post-traumatic stress disorder (PTSD), which was not associated with a neurotic predisposition, but was strongly correlated with horrific memories of the RTA. Mayou et al. (1993) also observed that the mental state three months after the RTA was highly predictive of the mental state one year after the accident. Additionally, they found that early information and psychological advice may reduce travel anxiety and distress, which may contribute to road safety. However, it was also shown that a too early psychological debriefing after an RTA may be less beneficial than a later psychological intervention (Mayou et al., 2000), whereas early and specific interventions such as solely addressing the phobic avoidance of travel can be more effective (Hobbs et al., 1996). Similarly, Taylor and Koch (1995) state, "motor vehicle accident PTSD can be effectively treated with systematic desensitization and in vivo exposure, often supplemented with other interventions such as cognitive restructuring. These treatments appear effective, even in cases of co-morbid chronic pain or neurological impairment" (pp. 733–734).

Today, desensitization and exposure can mean that a person decides to operate a vehicle again, making better experiences over time, which in turn may reduce the symptoms and the severity of PTSD. However, as yet it is unknown if this may be possible if vehicles are driving automatically.

In general, driving automation promises to increase road safety, which ultimately might also decrease the number and severity of RTA-related PTSD. Still, until fully automated vehicles (i.e, Level 5 driving, see SAE, 2018) are finally developed, human drivers will partly have to engage in vehicle operation. While partly automated driving systems are already available on the market (SAE Level 2, like the Tesla Autopilot or Cadillac Super Cruise), the next evolutionary step in this regard is conditional automation (SAE Level 3). Here, drivers may engage in Non-Driving Related Tasks (NDRTs), but they also need to remain fallback-ready. This means that they need to be able to respond safely to a Take-Over Request (TOR) issued by the driving system, e.g., when the system leaves its operational design domain due to sensor failures or external hazards like obstacles on the roads whose handling exceeds the system's capabilities. After receiving a TOR, the driver-passenger has to take-over control of the automated vehicle and resume manual driving in order to avoid an accident. This has strong implications for driving habits and experience. Drivers do no longer drive themselves for most of the time, but only in potentially critical situations. Manual driving experience may happen mostly in the context of TORs, which we argue, because of their potentially safety–critical aspects, could in turn be perceived differently by drivers who suffered a severe RTA.

Bennun and Bell (1999) distinguished between depression, anxiety, and PTSD symptoms and found that depression and anxiety often remit over time, whereas intrusion and avoidance do not always yield a remission. They further note that pre-accident predispositions may play an important role in developing psychological disorders after an RTA. Similar findings have been reported for driving anxiety and driving-related fear, which are not always a result of an RTA (Taylor et al., 2002). In a subsequent study, Taylor and Paki (2008) assessed the general prevalence of anxiety and fear related to driving. They identified that moderate to extreme anxiety about driving was described by 8% of all participants, and moderate to extreme driving fear was reported by 7% of the subjects.

Although psychological resilience has gained increased popularity in psychological research in the last few decades, to date, there exists no one exhaustive definition of this psychological construct. In contrast, several definitions for psychological resilience have been proposed, for example, the ability " … to adapt successfully in the face of stress and adversity." (Wu et al., 2013, p. 1) or " … to maintain or regain mental health, despite experiencing adversity." (Herrman et al., 2011, p. 258) or " … to overcome the difficulties experienced in the different areas of one's life with perseverance, as well as good awareness of oneself and one's own internal coherence by activating a personal growth project." (Sisto et al., 2019, p. 1) or " … a concept of healthy, adaptive, or integrated positive functioning over the passage of time in the aftermath of adversity." (Southwick et al., 2014, p. 1) or " … to adapt to the challenges of life and maintain mental health despite exposure to adversity" (Chmitorz et al., 2018, p. 78) among many others. However, severe RTAs are often experienced as adverse life events, which may also be mediated by psychological resilience.

Despite the deleterious effects of severe RTAs on mental health (mentioned above), to the best of our knowledge, there exists no SAE Level 3 driving simulator study which focuses on NDRT, TOR, and driving performance as well as on symptoms of PTSD, depression, *and* psychological resilience, both for participants who experienced an RTA vs. those who experienced no RTA. We believe this is particularly important though, since in SAE Level 3, drivers may engage in tasks other than driving (NDRTs) for the first time but still have to perform safety–critical tasks (TORs). Even if SAE Level 3 vehicles would expose drivers much less often to the danger of an RTA, driver's perception of and performance in SAE Level 3 driving may be impacted by the fact that they experienced a severe RTA in manual or assisted driving (SAE Levels 0–2).

1.1. Related studies on automated driving

Increasing SAE Levels of driving automation (e.g., from SAE Level 0 to 5; cf. SAE, 2018; Shuttleworth, 2019) are expected to enhance road safety. Koisaari et al. (2020) investigated current fatal at-fault crashes for SAE Level 2 (*partially* automated driving; i.e., adaptive cruise control (ACC) and lane keeping system (LKS)), and identified that active driver input until the crash, suicides, and difficult weather and hazardous road conditions as the most critical parameters for current active safety systems.

So far, there have been several studies on AVs which focused on different safety-critical aspects. In an experiment on different levels of blood alcohol concentrations (BACs: placebo, 0.05%, 0.08%) it was found that a BAC of 0.08% increases the TOR reaction time and negatively affects the longitudinal and lateral vehicle control, whereas 0.05% BAC resulted only in descriptive impairments (Wiedemann et al., 2018). In another study, Dixit et al. (2019) showed that risk attitudes and age have a significant impact on acceptance of AVs, the productivity in AVs (older persons are less productive), and safety (with risk averse individuals engaging slower). Similar studies on older drivers revealed that they benefited more from the longer notification interval between a TOR and the actual take-over than younger drivers, and that those who were more engaging in non-driving-related activities were braking harder (Clark & Feng, 2017). In line with these findings, but with a special emphasis on fully automated vehicles (FAV; SAE Level 5) it was found that the acceptance of FAV is high for older adults, especially if the FAV is perceived as reliable and operated in a similar driving style to their own (Haghzare et al., 2021).

Despite the advantages of AVs, their connectivity function makes them vulnerable to cyber-attacks either if the deviation between expected and measured behaviors exceeds a certain a priori set threshold, which would activate a security scheme, or if there might occur slight attacks below the threshold (Li et al., 2018). Furthermore, Liu et al. (2019) found that people judge RTAs with AVs more severely than those evoked by humans regardless of their severity (injury or fatality) and that a higher prior negative perception of AVs intensifies people's negative affect resulting in lower acceptance which may deter them from adopting AVs. Consistent with these findings, it was shown that attributions of causes of accidents in the media to *human error* are perceived as more likely and are more accepted than other causes such as a mechanical failure, a technical or a computer error (Nees et al., 2020).

Recently, Lin et al. (2020) simultaneously investigated individual risk differences in relation to NDRT, TOR, and driving performance. They compared three different conditions (i.e., no task, reading the news, and watching a video) for high crash risk (HCR) and lower crash risk (LCR) drivers in emergency take-over situations during a driving simulator study (SAE Level 3). Although in both groups the reaction times (e.g., brake reaction time, hands-on time) were faster for no task compared to the two other conditions, LCR drivers had shorter brake reaction times than HCR drivers, who themselves showed a lower hazard perception compared to the LCR drivers.

In a review on automated driving (AD), time budget, repeated exposure to take-overs, silent failures, and handheld secondary tasks were identified to significantly influence take-over time (McDonald et al., 2019). Additionally, the authors reported that post-take-over control is influenced by the driving environment, non-handheld secondary tasks, level of automation, trust, fatigue, and alcohol. Although they postulated models to explain these effects for future work, those models did not address the potential influence of severe RTA experiences.

To summarize, we found that there exist studies that investigate individual risk perception differences and various NDRTs on users in SAE Level 3 driving but none that investigates the impact of having experienced a severe RTA previously, that of utilized NDRT modality (i.e., visual vs. auditory) during AD, and potential interaction effects.

Thus, in the following, we present a driving simulator experiment involving subjects of severe RTAs that also varies the NDRT modality, to reveal potential interaction effects.

1.2. The present study

Given the ever-existing threat of an RTA and the fact that many people experienced an RTA of which some of them are suffering for a long time, it is crucial to investigate whether or not RTA experiences may affect NDRT, TOR, and driving performance. As it is well-documented that RTAs may have an influence on the manual driving behavior (Clapp et al., 2011; 2014; Mayou et al., 1993; 1997), which is critical for road safety, we were interested to specifically study whether or not experienced severe RTAs also may have an influence on the driving behavior in SAE Level 3 automated driving scenarios, which, in turn, would have an impact on traffic (or road) safety. Hitherto, there exists no driving simulator study which explores the potential influence of severe RTA experiences on NDRTs, take-over reactions, and driving performance in automated driving (SAE Level 3). Therefore, we investigated the following research questions (RQ) and hypotheses (H) in a medium-fidelity driving simulator:

RQ1: How is the influence of an experienced severe RTA on NDRT (e.g., a reading comprehension task using an auditory-speech vs. a heads-up display), TOR, and driving performance, compared to a control group with participants with no previous RTA experiences?

H1.1: In line with previous findings on detrimental psychological effects after an RTA and on arousal in the context of difficult tasks as described in the Yerkes-Dodson Law, we expect less good NDRT, TOR, and driving performance for those participants with severe RTA experiences compared to those who experienced no RTA. Additionally, we predict increased arousal (indicated by physiological stress and higher self-ratings of cognitive workload as well as lower ratings of trust) of those subjects with severe RTA experiences.

H1.2: Nevertheless, we do not assume a performance difference between the auditory-speech and the heads-up display.

Besides the investigation of our main research question, RQ1, we wanted to know whether we can replicate prior findings of severe RTA experiences on symptoms of PTSD and depression mentioned in the literature above. Additionally, we investigated psychological resilience in the context of potential deleterious effects of RTAs on mental health. The development of psychological resilience has been documented for numerous adverse life events and not necessarily only for severe RTAs (Chmitorz et al., 2018; Herrman et al., 2011; Sisto et al., 2019; Southwick et al., 2014; Wu et al., 2013). Hence, we expect no difference between the two groups (RTA vs. no RTA), because the participants of the no RTA group might have also experienced the development of psychological resilience (which we assessed).

RQ2: How is the impact of an experienced severe RTA on symptoms of PTSD, depression, and the potentially deleterious effects of stress, which may affect psychological resilience, in contrast to a control group whose participants experienced no RTA?

H2.1: We predict that participants with severe RTA experiences report more symptoms of PTSD and depression, as has been shown in earlier studies mentioned above.

H2.2: We hypothesize almost the same self-ratings of psychological resilience for both groups (RTA and no RTA) and assume that severe RTAs may not necessarily change resilience growth of the survivors when compared to the randomly selected participants of the no RTA group who might have also experienced the development of psychological resilience.

2. Method

2.1. Participants

Fifty-three fully-licensed drivers (27 female, 26 male) between 19 and 70 years old (female: M = 29.4, SD = 13.2; male: M = 29.0, SD = 12.5) were recruited to our driving simulator study via a newspaper advertisements (47.2 percent of all participants), via an institutional email sent out to all students and employees (24.5 percent), or by personal invitation (28.3 percent) at Technische Hochschule Ingolstadt. All of them were asked if they either experienced any *severe* accident within the last year or *no* car accidents at all. The period of one year was chosen because especially the ensuing year after an RTA is known that it may have significant consequences on the mental state and quality of life (Mayou & Bryant, 2002). Besides this, an RTA was defined as that a person has experienced a severe accident as a victim where she or he was innocent and experienced a severe threat that she or he could have potentially lost her or his life. Moreover, subjects who accidentally made a severe wrongdoer were also invited. Also in this case, severe was defined as a severe threat that she or he could have potentially lost her or his life. Twenty-one of them reported that they suffered from a slight and five of them from a severe physiological pain. Hence, 26 participants (12 female) were allocated to the *accident group* and 27 (15 female) to the *non-accident group*. Persons who experienced a *non-severe* or a *severe* car accident longer ago than one year were excluded from participation. All participants drove their car at least weekly and had a normal or corrected-to-normal vision. Moreover, all of them were fluent in German, consumed no alcohol or drugs, and reported no diagnosis of a psychiatric or neurological disorder. The study was conducted in accordance with the Declaration of Helsinki and was approved by the ethics committee of the Catholic University Eichstätt-Ingolstadt. External participants and students received 20 Euro, and employees of the local institution were allowed paid leave from their workplace but did not receive monetary compensation due to legal reasons.

Furthermore, 51 were from Germany, one from Spain, and one from Turkey. Twenty-six lived in a relationship, 11 were married, 13 were single, one was widowed, one divorced, and one did not specify the marital status. Twenty-eight of them lived *in* and 15 *near* a city, and nine of them in the countryside (one did not specify). Public transport was regularly used by nine of them. Eighteen used their bicycle regularly in spare time and 10 of them cycled regularly to work. Car- and/or ride-sharing was used by 10 participants.

2.2. Materials and equipment

In the following, we explain the self-rating scales (cf. 2.2.1 to 2.2.5), the driving and NDRT performance as well as the physiological measures assessed in the two driving simulator conditions: (1) Auditory-Speech-Display (ASD) and (2) Heads-Up Display (HUD; cf. 2.2.6). We deployed all questionnaires on LimeSurvey, Version 3.12.1 + 180616, (Limesurvey Project Team & Schmitz, 2021) and collected all data online and anonymously.

Cronbach's Alpha (Cronbach's α) is considered as a measure for internal consistency of a scale (Cronbach, 1951; Cronbach et al., 1963) which cannot be computed on a single item (e.g., the single-item-dimensions of the NASA-TLX), but across at least three or more items. We computed Cronbach's α for the following scales (observed in this study) and reported the values directly here accompanying the respective questionnaire, whereas values above 0.7 are considered as *acceptable* (Nunnally, 1978). However, it has to be noted that Cronbach's α is sensitive to the number of items and to the sample size such that more items and larger sample size are more likely to result in a higher value (Taber, 2018). Additionally, it depends on the intercorrelations and on the dimensionality of the items (Cortina, 1993).

2.2.1. Questionnaire: NASA-TLX

The NASA-TLX is a multi-dimensional scale developed to obtain workload estimates (Hart, 2006). It comprises a 6-factor structure assessing the six single-item-dimensions *Mental*, *Physical*, and *Temporal Demand*, as well as *Performance*, *Effort*, and *Frustration*. Participants responded to the six items using a 10-level rating-scale, whereas the items Mental (e.g., "How mentally demanding was the task?"), Physical (e.g., "How physically demanding was the task?"), and Temporal Demand, as well as Effort and Frustration are anchored at 0 (low) and 10 (high), and for Performance at 0 (good) and 10 (poor). Therefore, low values indicate low reported workload. The NASA-TLX can be analyzed for each single-item-dimension and as an overall *workload score* computed as sum score across the six items. We decided on focusing on the overall workload scores and identified a Cronbach α of 0.71 for the ASD and of 0.78 for the HUD condition.

2.2.2. Questionnaire: Trust scale

The Trust Scale (Jian et al., 1998; 2000) measures trust and distrust in technology and consists of 12 items comprising the secondary factors *Trust* (7 items; sample item: "I can trust the system") and *Distrust* (5 items; sample item: "I am suspicious of the system's intent, action, or outputs"). The items are rated on a 7-point scale anchored by "not at all to = 1" and "extremely = 7". For the factor Trust, we identified a Cronbach α of 0.89 for the ASD and 0.90 for the HUD condition. For the factor Distrust, we obtained a Cronbach α of 0.82 for the ASD and 0.77 for the HUD condition.

2.2.3. Questionnaire: PCL-5

The PCL-5 (Blevins et al., 2015; Weathers et al., 2013) is a 20-item self-report checklist for *posttraumatic stress disorder* (PTSD) according to the diagnostic criteria of the DSM-5 (American Psychiatric Association, 2013). The items are all positively coded (in terms of symptom severity) and provided with a 5-point Likert scale (0 = "Not at all" to 4 = "Extremely") constituting one overall factor summed as a *total severity score* ranging from 0 to 80 (sample item: "In the past month, how much were you been bothered by: Repeated, disturbing, and unwanted memories of the stressful experience?"). Based upon psychometric analyses, a cut-off score of 31–33 was suggested, whereas a total score of 31–33 or higher is considered that the patient may benefit from PTSD treatment. We identified a Cronbach α of 0.89 for this overall PTSD factor.

2.2.4. Questionnaire: PHQ-9

The *Patient Health Questionnaire* for *Depression* (PHQ-D or PHQ-9) covers nine positively coded items with a 4-point Likert scale anchored with "not at all = 0" and "nearly every day = 3", whereas each item represents one of the nine DSM-IV criteria for depression (Kroenke et al., 2001; Kroenke & Spitzer, 2002; Löwe et al., 2002). The nine items are summed to create one overall depression severity score ranging from 0 to 27 points (sample item: "Little interest or pleasure in doing things"). The total score is subdivided into 0–4 (*no depression*), 5–10 (*mild symptoms of depression*), whereas 10 is considered a cut-off point for a *major depression*. Hence, patients suffering from major depression report a value of 10 or greater, whereas 10–14 is interpreted as *medium*, 15–19 as *severe*, and 20–27 as *very severe* depression. The computation of Cronbach's α revealed a value of 0.85 for the total severity factor.

2.2.5. Questionnaire: RS-25

The Resilience Scale (Leppert et al., 2008; Schumacher et al., 2005; Wagnild & Young, 1993) measures the self-reported psychological resilience with 25 positively coded items and the two secondary factors *Personal Competence* and *Acceptance of Self and Life* on a 7-point rating scale ("Disagree = 1") to ("Agree = 7). The subscale Personal Competence encompasses characteristics such as self-confidence and perseverance (sample item: "When I make plans I follow through with them."), while the subscale Acceptance of Self and Life assesses characteristics such as adaptability and self-acceptance (sample item: "I usually take things in stride."). We obtained a Cronbach α of 0.87 for Personal Competence and 0.83 for Acceptance of Self and Life.

2.2.6. Driving simulator conditions and measures

A medium-fidelity movement-based (hexapod) driving simulator with a genuine VW Golf cockpit served as driving simulator (cf. Figs. 1 and 2). The cockpit hosted a simulated main front view, 2 side windows and a rear window. Added were a display in the cockpit's center stack showing the current driving mode (automated driving, manual driving) at all times and the TOR, when it was issued. Further modifications to the driving simulator are listed in the respective following sections.

Take-Over Task. We aimed at providing a potentially demanding scenario to investigate participants' responses to TORs. We designed a straight three-lane Highway with German standard layout, where the ego-vehicle was driving 120 km per hour in automated mode on the center lane. In front of the ego-vehicle, three lead vehicles (one per lane) were driving with the same speed. Besides the lead vehicles, there were no other vehicles in the scenery. When a TOR occurred, two of the three lead vehicles stopped, requiring the participant to maneuver the ego-vehicle in the remaining free lane. The TOR was issued multi-modal (white "Take-Over!" on red background as visual indication in parallel to an auditory "beep" signal) with a lead time of 5 s time-to-collision (i.e., not swerving would result in crashing with one of the stopped vehicles in 5 s). The center lane was always blocked to necessitate a reaction, and the remaining free lane was quasi-randomized (left/right lane). After clearing the situation, participants were told to maneuver the ego-vehicle back to the center lane, where the automated mode was activated again automatically after 2.5 s. The mode of the simulated system (automated, TOR, manual driving) was constantly displayed on a tablet computer mounted in the center stack.

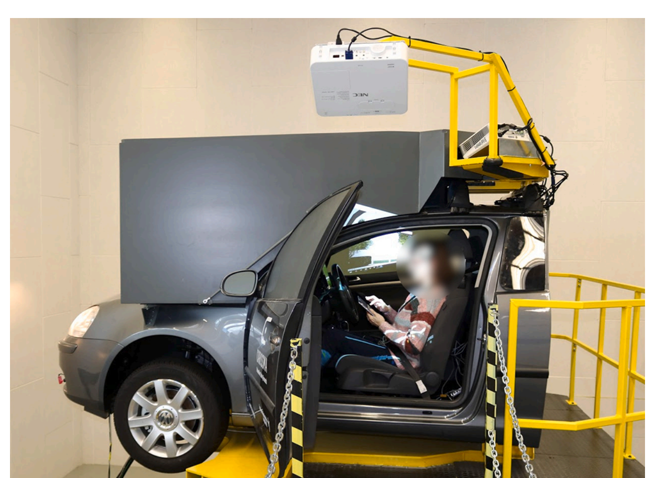

Fig. 1. The driving simulator with a participant completing post-condition questionnaires.

Fig. 2. Participant in the driving simulator during the HUD condition while engaging in the NDRT during automated driving.

Text Comprehension Task. As NDRT, we employed an adapted version of the reading-span task by Daneman and Carpenter (1980). Single sentences are continuously displayed with the respective condition's modality and have to be rated on their semantic correctness. Binary ratings were issued by the user with a computer mouse in the driving simulator's middle console that has its left button marked green (sentence is semantically correct) and its right button marked red (sentence is semantically incorrect). After a rating was issued, feedback was issued to the user either visually (green check mark or red cross) or via text-to-speech ("correct" or "incorrect"). If the user did not rate a sentence within 10 s, it timed out (cartoon-snail / "too slow" as feedback). A TOR interrupting the task led to discarding the sentence. In total, ca. 80 sentences were issued per condition. The involved *text perception and comprehension* acts as a basic (but comparable and measurable) (sub-) task of many desired future NDRTs in automated vehicles, such as reading, social media or office work (Pfleging et al., 2016). Since these activities potentially represent a significant threat to safety, experts argue a need for specialized in-vehicle interfaces to support them (Kun et al., 2016). Auditory and Heads-up Displays are, already today, somewhat widespread, accepted and well-researched (Bazilinskyy and Winter, 2015; Pauzie, 2015). Due to their beneficial properties in regards to road safety, we expect them to be extensively used for NDRTs -however, they are still inherently different considering their utilization of different attentional resources (Wickens, 2002). If, for example, persons who experienced a severe RTA, can also use a potentially visually distracting HUD or an auditory display that may overlap with auditory warning signals and maintain safety, productivity, and low stress levels, is unclear. Hence, we investigate (visual) HUDs and Auditory-Speech-Displays (ASDs) in conjunction with the text comprehension task as within-subjects conditions.

Auditory Speech Display condition. In this condition, sentences of the text comprehension tasks and feedback to their ratings were issued verbally by the free Google Text-To-Speech Engine. Prior to the conditions, participants chose between 7 voice output speed levels ($0.5\times - 3.0\times$, 0.5 difference per step) based on a sample sentence played with the respective speeds.

Heads-up display condition. During the condition, sentences and rating feedback was displayed visually on a simulated HUD in the top-third field of view of participants facing the road center (cf. Fig. 2). According to Tsimhoni et al. (2001), this HUD positioning is favorable in terms of workload on the user. The HUD was realized as an overlay to the driving simulator visualization with white background (50% transparency) and black font. Participants' vision was normal or corrected-to-normal to ensure no differences in visual reading ability.

Take-Over Reaction measures. The reaction time to the first driving action (RTd, seconds) was calculated as the time between the TOR was issued and the first input by the driver, i.e., a steering wheel change $> 2^\circ$ or a pedal actuation $> 5\%$ (Wintersberger et al., 2017). As second temporal measurement, we utilized the reaction time until the first gaze on the road (RTe, seconds), i.e., the time between issuing the TOR until the gaze of the driver enters the road area of interest (determined and calculated by the iMotions eye tracking software).

Driving Performance measures. To quantify participants' performance in the driving task following the TOR response, we calculated three parameters. The first is the time to collision on lane change (TTCoLC, seconds), given by the vehicles' center of mass entering the lane divider (i.e., the higher this value, the faster the reaction). We also calculated the steering reversal rate (SRR, i.e., the number of times per minute the steering actions' direction changes; a measure to quantify lateral driving stability; cf. SAE J2944, 2015), whereas higher values represent more steering input, i.e., worse lateral driving performance, as given by (Knappe et al., 2007). Finally, we also determined the average brake actuation (Brake Act., in %; relative to the pedal's maximum actuation), where higher values indicate stronger braking, for example due to shock reaction in a perceived emergency situation.

NDRT Performance measures. We calculated the average time from the start of displaying a sentence until the user rates it (RTt, seconds) to evaluate speed. To evaluate correctness, we utilized the F1-Score which balances the influence of task precision and recall on accuracy via the harmonic mean (F1-Score $= \frac{TruePositives}{TruePositives+0.5*(FalsePositives+FalseNegatives)}$).

Physiological measures. For physiological measurements, we primarily utilized a 500 Hz g.tec Nautilus system that synchronously recorded 3-electrode electrocardiography (ECG) and skin conductance with two electrodes. ECG electrodes were placed under the left and right clavicle near the left and right shoulders, and on the lower edge of the left rib cage. Skin conductance electrodes were prepared with electrode gel and attached to the volar middle phalanges of the participant's left hand's middle and ring fingers. For eye-tracking and related measures, we utilized the head-mounted Tobii Pro Glasses 2. We analyzed area-of-interest based glance measures that are not presented in this article as they are beyond its scope (cf. Declaration of Competing Interests). The open toolbox Ledalab for Matlab (Benedek & Kaernbach, 2014) was utilized for preprocessing raw and calculating skin conductance level (SCL) and galvanic skin response (GSR) measures. HRVTool (Vollmer, 2019) was used for analyzing heart rate and heart rate variability (HRV).

Electrodermal activity, i.e., measured skin conductance, as well as pupil diameter, heart rate, and heart rate variability were previously linked to arousal, workload and similar concepts (Stanton et al., 2004), specifically also in driving situations (Brookhuis et al., 2003; Brown & Huffman, 1972; Collet et al., 2009). In particular, we calculated the average skin conductance level (SCL), normalized pupil diameter (PupilDia) and HRV parameters during the drives per condition as measures of tonic (i.e., longer-term) workload. HRV parameters calculated include the heart rate (HR), the root mean square of successive heartbeat differences (RMSSD), the standard deviation of normal to normal RR intervals (SDNN), and the ratio of low- and high-frequency power (LFHFR), covering a range of widely used HRV indicators (Wang & Huang, 2012).

To investigate immediate stress (phasic workload) induced by TORs, we analyzed the phasic portion of the skin conductance level, also called the "Galvanic Skin Response" (GSR) and indicated by short-term event-related spikes in skin conductivity (termed Skin Conductance Responses, SCRs). The analysis time window ranged from second 1 to 11 after a TOR was issued. Utilizing Ledalab and its built-in Decomposition Analysis (Benedek & Kaernbach, 2010), we calculated the average phasic activation during significant SCRs (SCR, in μS), the number of SCRs (nSCR), and the maximum phasic activation (Ph.Max. in μS).

2.3. Design

We adopted a quasi-experimental two-factorial 2 × 2 design, with the between-subjects-factor *accident* (accident vs. non-accident group; quasi-experimental variable) and the within-subjects-factor *NDRT modality* (ASD vs. HUD) as independent variables, respectively. Hence, each participant experienced the ASD as well as the HUD condition (cf. 2.2.6). The assignment to those conditions was counterbalanced across all participants (i.e., the participant with ID = 1 started with the ASD condition followed by HUD, and the participant with ID = 2 began with the HUD condition followed by ASD, and so forth). Our dependent variables were measured by the aggregated sum- or mean scores of the aforementioned psychometric questionnaires (cf. 2.2.1 to 2.2.5) and by the TOR, the driving performance, the NDRT, and the physiological measures (cf. 2.2.6).

2.4. Procedure

After welcoming the participants, they were introduced to the driving simulator study (cf. the workflow of the study below in Fig. 3). Then, all participants were shown the driving simulator (cf. Fig. 1) and were asked, 'if they still feel good to participate in our study or not'. Hence, everyone could explicitly redecide if she or he wants to take part in the study or not. However, everyone decided to continue and provided written informed consent prior to the beginning of the study. Then, we introduced the participants to all parts of our study. In the first part of the study, the questionnaires on mental and physical health (PCL-5, PHQ-9, RS-25) accompanied by demographic questions had to be filled in (cf. 2.2.3 to 2.2.5). Then, the driving simulator was shown again to all participants, and they were asked again, 'if they still feel good to participate in our study or not?' All of them decided to continue.

In the second part of the study, the necessary tools and electrodes for the physiological measures (cf. 2.2.6) had to be put on. Then, each participant was introduced to the driving simulator and had to complete a short training session (about 5 min with two take-overs without a NDRT) to make themselves familiar with it and learn how to perform a take-over reaction.

Participants were told that the driving automation system can reliably perform the dynamic driving task until a TOR, which is issued due to some event that is beyond the capabilities of the vehicle or a sudden system failure, and that they will have to respond timely in case of such a TOR.

Once the participant said that she or he is familiar with the whole setup, the first driving simulation condition started (either ASD or HUD; counterbalanced as described in 2.3). One driving simulator condition took around 12 min and included 4 "automated driving with NDRT, interrupted by take-over and handback" phases. Each condition started with a 20-seconds baseline recording of pupil diameter after the simulator door was closed and the driving simulation loaded. Then, the simulation started in automated driving mode with the first NDRT sentence. After each of the two conditions, everyone had to complete the state scales such as the NASA TLX and the Trust Scale (for within-subjects comparisons; cf. 2.2.1 and 2.2.2) prior to continuing with the second driving simulation condition. The body-worn devices and electrodes for the physiological measures were removed after the second condition.

Finally, everyone was thanked, debriefed, and paid (cf. 2.1). Additionally, everyone was offered a "Take-Home-Sheet" with all important information about the study and contact details of the study examiner in case questions may arise later on. Before they left, everyone had to sign a form that they felt capable of driving a car, cycling, or participating in public transport (if they actually felt capable), and that they will not participate in road traffic until 30 min after the last driving simulation condition. The entire study duration ranged from 90 to 120 min. All participants completed the study.

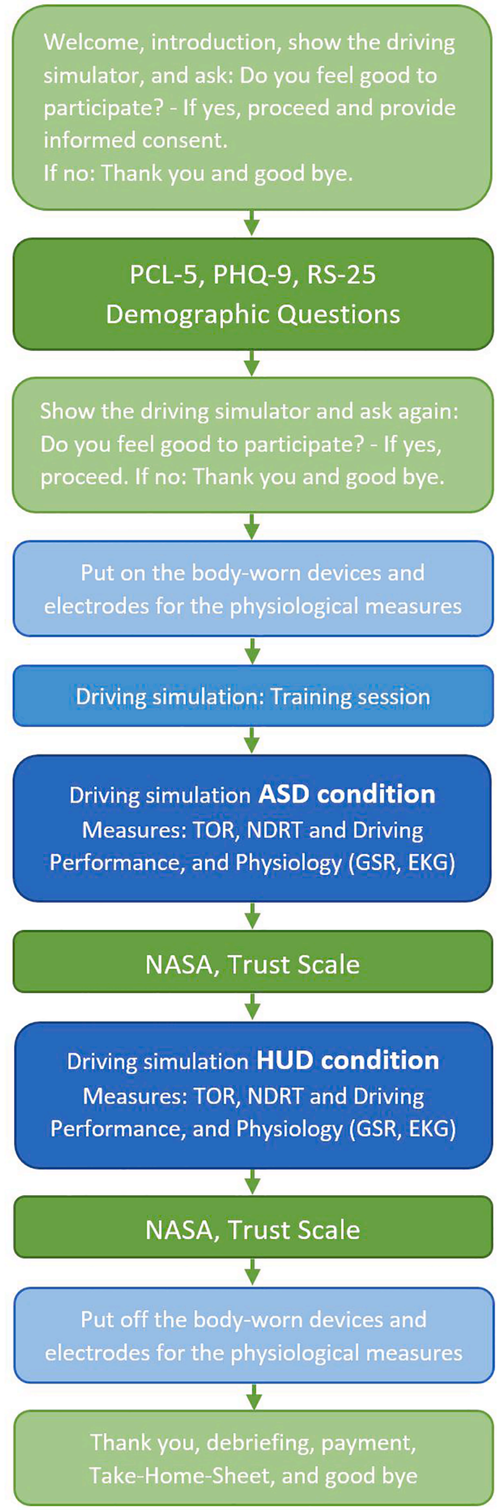

Fig. 3. The workflow of the driving simulator study. The ASD and the HUD condition (dark blue background color) were counterbalanced across all participants. The questionnaire parts (dark green background color) and all other parts remained unchanged during the study. Administrative tasks are depicted in light green.

2.5. Statistical analyses

We applied a two-step procedure for our statistical analyses. First, we performed several repeated measures multivariate analyses of variances (MANOVAs) to investigate the influence of the independent variables

(IVs) *accident* (accident vs. non-accident group) and *NDRT modality* (ASD vs. HUD; *repeated measures*), each on a specific bundle of content-related dependent variables (DVs) such as TOR, driving, and NDRT performance, cognitive workload, and physiological measures such as GSR and ECG, and the secondary factors of the scales such as the NASA TLX and the Trust Scale. The advantage of a MANOVA is to simultaneously investigate the influence of IVs on a set of DVs, while controlling for the family-wise Type I error rate across all comparisons to avoid alpha cumulation due to multiple testing. However, the disadvantage is that in case of a significant effect of a content-related variable bundle, it is not always necessarily clear from which variable(s) the overall significance is evoked.

Therefore, and second, we applied repeated measures univariate analyses of variances (ANOVAs) within each bundle of content-related DVs to investigate the data in more detail. Although the univariate analyses would have been only necessary for the significant variable bundles, for the sake of completeness and transparency, we provide all results in Table 1. Nevertheless, as common, we only interpreted the significant ANOVAs.

The statistical assumptions such as normality and variance homogeneity have been met in most conditions. Additionally, we investigated the distributional assumptions with QQ-Plots, skewness, and kurtosis which confirmed that repeated measures (M)ANOVAs could be applied on the almost equal cell sample sizes (26 accident; 27 non-accident). We applied Pillai's trace for the MANOVA, which is more robust if the statistical assumptions are not met (Ateş et al., 2019; Pillai, 1955; Seber, 1984). On top of this, we investigated if the DVs within one variable bundle were not correlated too high. We observed that all of them were correlated below 0.8, most of them below 0.7, which are good results (clearly below 0.9).

Additionally, we performed Mann-Whitney-U-Tests on the secondary factors of the questionnaires of the PCL-5, the PHQ-9, and the RS-25 to investigate whether or not participants who experienced a severe traffic accident differ or not from those who had no accident experience (Note: Additional analyses with the independent two samples t-Test yielded the same results, but are not listed in the Results section). Since these questionnaires have only been completed *once* by the participants (in contrast to the NASA TLX and the Trust Scale after each condition; cf. Fig. 3) no repeated measures MANOVAs have been applied. The *overall* significance level was set at $\alpha = 0.05$, whereby smaller p values were considered as significant. To control the Type I error rate because of multiple testing, the Bonferroni-Holm correction method (Holm, 1979) was applied, which is less conservative than the Bonferroni correction and also adjusts for the Type II error rate yielding more statistical power. For all results, appropriate effect size measures have been reported.

All statistical analyses have been performed with IBM® SPSS® Statistics, Version 25 (IBM Corp., 2017).

3. Results

3.1. Driving simulator results: ASD vs. HUD for the accident and the non-accident group

First, we studied the possible influence of experienced severe road traffic accidents (Yes vs. No) on take-over reactions and NDRTs when using an ASD and a HUD in automated driving (RQ1). Therefore, we tested the following driving and NDRT performance, physiological, and psychological measures:

TOR measures. Although we found no group difference for the between-subjects variable *accident* (Yes vs. No), F (2, 45) = 0.97, p = 0.385, $\eta^2_p = 0.04$, we identified clear differences for the within-subjects variable NDRT modality (ASD vs. HUD), F (2, 45) = 11.87, p = 0.000, $\eta^2_p = 0.35$, on the DVs. Thereby the time to first driving action (RTd) was faster for HUD than for ASD. In contrast, the time to first gaze on the road (RTe) was faster for the ASD and slower for HUD condition (cf. Tables 1 and 2). However, there was no interaction between accident and NDRT modality, F (2, 45) = 1.69, p = 0.197, $\eta^2_p = 0.07$.

Driving performance. We could not uncover any substantial differences neither for the grouping variable accident, F (3, 49) = 2.07, p = 0.116, $\eta^2_p = 0.11$, nor for the repeated-measures variable NDRT modality, F (3, 49) = 1.58, p = 0.205, $\eta^2_p = 0.09$, or the interaction between both IVs, F (3, 49) = 0.87, p = 0.463, $\eta^2_p = 0.05$, on either of the DVs (cf. Tables 1 and 2).

NDRT performance. The results indicated no group difference for the

Table 1
Two-way Univariate Analyses of Variances of the independent variables ASD vs. HUD and accident vs. non-accident on the dependent variables (DVs) of TOR, driving, and NDRT performance, the physiological measures, and the secondary factors of the NASA TLX and the Trust Scale (TS).

			Univariate								
			Accident vs. Non[a]			ASD vs. HUD			Interaction[b]		
	DVs[c]	*df*	*F*	*p*	η^2_p	*F*	*p*	η^2_p	*F*	*p*	η^2_p
TOR	RTd	1, 46	0.02	0.903	0.00	**5.25**	**0.027**	**0.10**	1.86	0.180	0.04
	RTe		1.99	0.166	0.04	**15.25**	**0.000**	**0.25**	2.18	0.147	0.05
Driving	TTCoLC	1, 51	0.97	0.329	0.02	1.12	0.296	0.02	0.25	0.618	0.01
	SRR		0.78	0.382	0.02	3.43	0.070	0.06	2.09	0.154	0.04
	Brake Act.		4.44	.040[d]	0.08	1.31	0.257	0.03	0.63	0.430	0.01
NDRT	RTt	1, 51	0.65	0.423	0.01	**17.40**	**0.000**	**0.25**	0.11	0.742	0.00
	F1 Score		0.03	0.870	0.00	0.08	0.778	0.00	1.36	0.249	0.03
Workload	SCL	1, 46	0.28	0.599	0.01	2.81	0.101	0.06	0.002	0.965	0.00
	Pupil D.		**14.08**	**0.000**	**0.23**	**31.61**	**0.000**	**0.41**	1.29	0.261	0.03
GSR	SCR	1, 51	1.37	0.248	0.03	0.01	0.929	0.00	0.00	0.989	0.00
	nSCR		0.02	0.888	0.00	0.004	0.947	0.00	1.69	0.199	0.03
	Ph. Max.		0.48	0.493	0.01	0.70	0.406	0.01	0.97	0.330	0.02
ECG	HR	1, 43	0.46	0.502	0.01	1.73	0.195	0.04	0.21	0.646	0.01
	RMSSD		2.78	0.103	0.06	0.06	0.805	0.00	0.34	0.562	0.01
	SDNN		0.86	0.359	0.02	0.10	0.749	0.00	0.22	0.643	0.01
	LFHFR		0.02	0.879	0.00	0.06	0.807	0.00	0.49	0.490	0.01
Scales	NASA TLX	1, 46	0.51	0.477	0.01	**12.62**	**0.001**	**0.22**	3.97	0.052	0.08
	TS: Trust		1.98	0.166	0.04	0.52	0.473	0.01	4.26	0.045	.09[d]
	TS: Distrust		0.10	0.748	0.00	0.02	0.894	0.00	0.16	0.690	0.00

Note. N = 53. η^2_p denotes the *partial* η^2; [a]Non = Non-Accident group; [b]Interaction between Accident vs. Non-Accident and ASD vs. HUD; [c]All DVs are explained in section 2.2.6; [d]Neither multivariate analyses, nor Bonferroni-corrected ANOVAs are significant; Significant results are highlighted in boldface.

Table 2
Mean and standard deviation of driving and NDRT performance, the physiological measures, and the secondary factors of the NASA TLX and the Trust Scale (TS) as a function of the accident and NDRT modality.

		Accident				No accident			
		ASD		*HUD*		*ASD*		*HUD*	
	DVs[a]	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
TOR	RTd (sec.)	**2.61**	**0.48**	**2.53**	**0.58**	**2.67**	**0.59**	**2.48**	**0.55**
	RTe (sec.)	**0.11**	**0.20**	**0.30**	**0.22**	**0.10**	**0.17**	**0.18**	**0.17**
Driving	TTCoLC (sec.)	2.77	0.41	2.61	0.53	2.85	0.65	2.79	0.77
	SRR (r/min.)	62.63	11.90	63.11	11.85	58.35	11.70	62.25	10.30
	Brake Act. (norm.)	0.02	0.03	0.02	0.03	0.04	0.04	0.04	0.03
NDRT	RTt (sec.)	**5.44**	**0.65**	**4.92**	**0.94**	**5.34**	**0.53**	**4.74**	**1.03**
	F1 Score	0.47	0.10	0.48	0.11	0.49	0.08	0.46	0.08
Workload	SCL (μS)	10.82	5.67	10.39	5.77	10.94	5.61	10.25	4.56
	Pupil D. (norm.)	**0.98**	**0.05**	**0.91**	**0.06**	**1.02**	**0.07**	**0.97**	**0.05**
GSR	SCR (μS)	1.94	1.24	1.94	1.04	1.62	0.91	1.63	0.95
	nSCR (n)	7.57	1.64	7.18	1.55	7.22	2.28	7.65	2.09
	Ph. Max. (μS)	6.33	3.07	6.29	3.02	5.54	2.21	6.06	3.14
ECG	HR (bpm)	78.25	12.74	78.44	14.49	79.55	8.95	80.61	10.32
	RMSSD	39.46	22.49	41.98	21.44	33.20	12.40	33.65	13.75
	SDNN	58.98	18.39	60.01	21.44	55.84	18.29	56.13	17.02
	LFHFR	3.41	2.21	3.62	2.92	3.53	2.50	3.34	1.84
Scales	NASA TLX	**25.04**	**6.43**	**26.19**	**5.82**	**25.15**	**8.70**	**28.81**	**10.46**
	TS: Trust	26.84	7.67	28.31	8.29	27.60	5.63	25.32	6.52
	TS: Distrust	7.88	5.17	7.85	5.03	7.27	4.92	7.62	5.00

Note. N = 53. [a]All DVs are explained in section 2.2.6; M and SD are highlighted in boldface for significant group differences between ASD vs. HUD; There are no significant group differences between accident and no accident (cf. Table 1).

variable accident, F (2, 50) = 0.41, p = 0.668, η^2_p = 0.02, on both NDRT measures, whereas we could reveal a sufficiently large effect for NDRT modality, F (2, 50) = 8.85, p = 0.001, η^2_p = 0.26, the average task response time (RTt) in favor for the HUD condition enabling faster responses than for the ASD condition (cf. Tables 1 and 2). However, no interaction effect was found between both IVs, F (2, 50) = 0.78, p = 0.464, η^2_p = 0.03 (cf. Table 1).

Workload measures. We found an overall group difference for the variable accident, F (2, 45) = 7.29, p = 0.002, η^2_p = 0.25, whereby the univariate analyses were only uncovered a substantial difference for pupil diameter in favor for the accident group with a smaller pupil diameter than the non-accident group (cf. Tables 1 and 2). Additionally, there was a significant difference for NDRT modality, F (2, 45) = 16.82, p = 0.000, η^2_p = 0.43, in favor for the HUD condition with a smaller pupil diameter than for the ASD condition. Still, there was no interaction effect between both IVs, F (2, 45) = 0.63, p = 0.535, η^2_p = 0.03.

GSR (stress) measures. The results revealed no meaningful difference neither for the grouping variable accident, F (3, 49) = 0.63, p = 0.601, η^2_p = 0.04, nor the repeated measures variable NDRT modality, F (3, 49) = 0.31, p = 0.822, η^2_p = 0.02, or the interaction between both independent variables, F (3, 49) = 0.73, p = 0.540, η^2_p = 0.04, on either of the DVs (cf. Table 1).

ECG measures. There were no differences among any of the measures of the electrocardiogram neither for the two IVs accident, F (4, 40) = 1.66, p = 0.178, η^2_p = 0.14, and

NDRT modality, F (4, 40) = 1.85, p = 0.138, η^2_p = 0.16, nor the interaction between both IVs,

F (4, 40) = 0.31, p = 0.871, η^2_p = 0.03 (cf. Tables 1 and 2).

Scales. Albeit there was no group difference for the variable accident, F (3, 44) = 1.70, p = 0.180, η^2_p = 0.10, we uncovered a significant difference for NDRT modality, F (3, 44) = 4.06, p = 0.013, η^2_p = 0.22, for the overall NASA TLX score indicating lower reported cognitive workload for the ASD condition compared to the HUD condition. However, there were no differences for the factors Trust and Distrust of the Trust Scale among the NDRT modalities. Again, no interaction effect between both IVs was found, F (3, 44) = 2.28, p = 0.093, η^2_p = 0.13 (cf. Tables 1 and 2).

3.2. PTSD, depression, and resilience ratings for the accident vs. The non-accident group

Second, we studied the potential impact of experienced severe road traffic accidents on self-reported symptoms of PTSD, depression, and potentially deleterious effects of stress which could affect psychological resilience, compared to the control group whose participants experienced no road traffic accidents.

The results of the Mann-Whitney-U-Tests on the independent grouping variable accident (Yes vs. No) on the total score of the PCL-5 questionnaire (i.e., the PTSD severity score) revealed a grouping difference with a *medium* to a *large* effect size of r = 0.4 according to Cohen (1992); An effect size of r = 0.3 or 0.5, indicates a medium or a large effect, respectively). Hence, participants who experienced an accident report significantly larger PTSD values than those who experienced no accident (cf. Table 3). Nevertheless, it has to be noted that besides this statistically large effect, even the Median of 9 for the accident group is clearly below the suggested cut-off score range of 31 to 33. Moreover, only one participant with a total severity score of 46 was above this cut-off score range, whereas all others were below (cf. 2.2.3). Although we excluded the participant with the total severity score of 46 in supplementary analyses, the group difference remained, Z = -2.75, p = 0.006, r = 0.38, with the same Medians. However, there was no difference for the

Table 3
Results of the Mann-Whitney-U-Tests for the independent variable accident on the secondary factors of the questionnaires PCL-5, PHQ-9, and RS-25.

	Accident				
	Yes	No			
Faktor	*Mdn*	*Mdn*	*Z*	*p*	*r*[a]
PCL-5: Severity score	**9.0**	**4.0**	**−2.91**	**0.004**	**0.40**
PHQ-9: Severity score	2.0	4.0	−1.55	0.121	0.21
RS-25: Personal competence	98.0	96.0	−0.26	0.796	0.04
RS-25: Acceptance of self and life	45.5	42.0	−0.79	0.433	0.11

Note. N = 53; Mdn = Median; [a]Effect size according to Cohen (1992). Significant results are highlighted in boldface.

grouping variable accident neither on the PHQ-9 depression severity score, nor on the two resilience scale factors of the RS-25, such as Personal Competence and Acceptance of Self and Life (cf. Table 3).

4. Discussion

The present automated driving simulator study (SAE Level 3) was designed to investigate if severe RTA experiences may influence the NDRT, TOR, and driving performance, physiological measures, and self-reported cognitive workload and trust in AVs within the two NDRT conditions ASD and HUD, respectively. Additionally, we assessed general symptoms of PTSD and depression as well as self-ratings of the two resilience subscales Personal Competence and Acceptance of Self and Life.

4.1. NDRT: ASD vs. HUD condition for the accident and the non-accident group

The main goal of the present study was to compare the accident and the non-accident group in the ASD and HUD condition (between-subjects comparison), both NDRT conditions with each other across all subjects (within-subjects comparison), as well as a potential interaction between both IVs (i.e., a 2 × 2 design; RQ1) on all outcome measures of NDRT, TOR, driving performance, physiology, cognitive workload, and trust. Although we could reveal a difference for pupil diameter in favor for the accident group with a smaller pupil diameter than the non-accident group, we observed no significant difference between the accident and the non-accident group for all other 18 DVs (cf. Table 1). This is the most striking result to emerge from our empirical data at SAE Level 3, because it could not necessarily be expected based on previous findings on detrimental psychological effects after an RTA at SAE Level 0 to 2 (Bennun & Bell, 1999; Frank, 1993; Kuch, 1993; Kuch et al., 1991; 1994; Mayou et al., 1993; 2000; Mayou & Bryant, 2002).

Moreover, it could be interpreted as a positive result for those driver-passengers with severe RTA experiences to not differ from those with no RTA experiences in important measures such as NDRT, TOR, driving performance, physiology, and trust, except for cognitive workload, but only for pupil diameter. Thereby, people with RTA experiences had a smaller pupil diameter than the non-accident group, which may indicate slightly better values for the RTA group because having a larger pupil diameter is usually interpreted as indicator for more alertness (Beggiato et al., 2018; Rendon-Velez et al., 2016; Tseng et al., 2018). Perhaps they feel a bit more secure in an AV than the subjects of the non-RTA group. However, this single significant outcome measure has to be considered within the context of the other 18 non-significant outcome measures (cf. Tables 1 and 2), indicating no group differences. Hence, the present findings should neither be interpreted that people with RTA experiences are performing better or are trusting more in automated driving, nor should it be used as a justification that people with RTA experiences perform as good as people with no RTA experiences.

In terms of our second IV NDRT modality, we uncovered a significant difference between the ASD and the HUD conditions only for five out of 19 outcome measures. Although the subjects did not differ between both NDRT conditions for 14 DVs, it is remarkable that the participants performed better in the HUD condition for the DVs *time to first driving action* (RTd), *the average task response time* (RTt), with a smaller *pupil diameter* than in the ASD condition. Interestingly, they achieved a shorter time to first gaze on the road (RTe) and reported less cognitive workload (NASA TLX) in the ASD condition. Especially shorter average RTe may have been caused due to the ASD simply requiring no visual attention for the NDRT. Hence we assume that participants generally used this freedom to anticipate a TOR by looking at the road already during the NDRT.

However, we observed no interaction effect between both IVs (cf. Table 1). Hence, the marginal mean differences between the accident and the non-accident group depicted in Table 2 did not significantly influence the DVs of the ASD and the HUD condition.

Albeit we cannot accept H1.1, we believe that these results positively contribute to the existing body of research in two ways. First, as RTAs are, as yet, an ever-existing threat to all road users, it is important to know that accidents may not necessarily influence driving, TOR, or NDRT performance, nor other physiological stress measures, or self-ratings of cognitive load or trust at SAE Level 3. This is especially surprising, because in contrast to our predictions that subjects who experienced a severe RTA should show very high arousal, which should result in a poorer performance according to the Yerkes-Dodson Law (Yerkes & Dodson, 1908), they actually did not show, nor report a higher arousal during NDRT and TOR tasks than those with no RTA experiences. This finding is still in line and can be even explained with the Yerkes-Dodson Law, because it may indicate that participants who experienced a severe RTA do not experience a too high physiological or cognitive arousal. Hence, they do not necessarily experience detrimental effects on NDRT or TOR performance. Second, taken together all the NDRT findings (14 non-significant and five significant), they indicate slightly better performance for the HUD condition, whilst the participants subjectively experience more cognitive workload than in the ASD condition. Hence, we can only partly accept H1.2.

4.2. PTSD, depression, and resilience ratings for the accident vs. The non-accident group

The second goal of the present study was to assess symptoms of PTSD and depression in self-report questionnaires and the self-ratings on the two resilience subscales Personal Competence and Acceptance of Self and Life for the accident and the non-accident group, respectively (RQ2). As predicted, participants with severe RTA experiences reported more symptoms of PTSD compared to those who had no RTA (H2.1). This finding is in line with previous results (Bennun & Bell, 1999; Mayou et al., 1993; Mayou & Bryant, 2002; Murray et al., 2002). However, it has to be mentioned that only one participant with a total severity score of 46 was above the suggested cut-off score range of 31 to 33, whereas all other scores were below 31. This indicates that except the one person, all others may not benefit from a PTSD treatment as proposed by the developers of the PCL-5 (Blevins et al., 2015; Weathers et al., 2013), although their values are subclinical.

Interestingly, and in contrast to our predictions on self-reported symptoms of depression, our results did not reveal a difference between the two groups (RTA vs. no RTA). This finding is also not strictly consistent with previous literature, where there are reported more symptoms and greater depression scores after an RTA (Bennun & Bell, 1999; Mayou et al., 1993; Mayou & Bryant, 2002). Nevertheless, Bennun and Bell (1999) also found that the ratings of depression decrease the longer the delay is between the RTA and the assessment. Although all our participants of the accident group reported that they experienced a severe RTA within the last year, most of them told us that it was already several months ago. Hence, this result seems realistic. Thus, we can partly accept H2.1 for self-reported symptoms of PTSD, but not for self-ratings of depression.

However, there was no difference for the grouping variable accident on either of the two resilience scale factors Personal Competence and Acceptance of Self and Life of the RS-25. We assume that severe RTAs may not necessarily change psychological resilience growth compared to other life events experienced by both groups. Hence, we can accept H2.2.

4.3. Limitations and future research

As there are many avenues for future accident research on SAE Level 3, we want to note that perhaps the tasks in both NDRT conditions might have been too easy for the participants such that they possibly did not differentiate enough between the accident and the non-accident group. Additionally, it would be interesting how our results may change if the delay between the severe RTA and the driving simulator study may be shorter, for example, one to three months. Nevertheless, we want to stress that besides the necessary ethical considerations and the crucial and potential problems of a re-traumatization and adverse psychological effects, as have been observed after a too early debriefing (Hobbs et al., 1996; Mayou et al., 2000), it would not be so easy to recruit probands for such a study, as a larger time gap may attract more participants. Furthermore, we want to mention that it might have been interesting to additionally assess whether the severe RTA was caused by a wrongdoer or experienced by the participant as an innocent victim. The cause for the accident might have an influence on the mental state of the participant which, in turn, could affect the outcome measures.

Considering technical limitations, a driving simulator does not provide the same level of realism and thus may diminish perceived danger. However, one might argue that when SAE Level 3 vehicles are employed and widely owned by people, participants may also not react as harshly to take-over situations, in comparison to when first experiencing such situations for the very first time.

Although we identified a null result that participants who experienced a severe RTA did not differ from those with no RTA experiences in almost all measures applied in our study (described above), we want to note that the overall sample size of N = 53 is not sufficiently large enough to generalize this finding. Henceforward, future studies with larger sample sizes, more sophisticated driving simulators as well as real (non-simulator) automated driving studies will be necessary, to shed more light on this crucial between-group comparison (i.e., RTA vs. no RTA) in the context of driving automation at SAE Level 3.

5. Conclusion

Currently, RTAs are an omnipresent threat and a common danger to all road users. Everyone faces the high personal, societal, and economic costs of RTAs to a smaller or larger extent, which will not be automatically eliminated when the first AVs will appear in the near future. Hence, the present driving simulator study is the first step towards a better understanding of potential influences of severe RTAs on two different scenarios involving NDRT engagement during automated driving (i.e., with ASD or HUD) and occasional safety-critical TORs. The most important result is that even severe RTA experiences do not undermine NDRT, TOR, and driving performance. However, this evidence should not be used as a justification that severe RTA experiences may not play a crucial role when designing or releasing AVs. In contrast, we want to stress that these findings need to be interpreted with caution and even more future studies on potential influences of RTA experiences are necessary. Despite these results, we observed a better performance for the HUD, whereas a lower cognitive workload was reported for the ASD across both groups, respectively. Interestingly, although the ASD is perceived with less cognitive workload, the HUD seems to enable better interaction performances in the NDRT condition. Albeit technology trust, depression, and psychological resilience ratings do not differ for both groups, as expected, we found higher self-ratings of PTSD symptoms for the accident than for the non-accident group. Hence, we agree with previous findings that psychological treatment at the right time after a severe RTA may reduce the psychological suffering and may be beneficial for mental health.

CRediT authorship contribution statement

Klemens Weigl: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft. **Clemens Schartmüller:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. **Philipp Wintersperger:** Conceptualization, Investigation, Methodology, Resources, Visualization, Writing – original draft. **Marco Steinhauser:** Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing. **Andreas Riener:** Conceptualization, Methodology, Funding acquisition, Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

Interim findings of the first 32 participants (and not of the full sample N = 53), but

(i) only focusing on the comparison of the two NDRT conditions (i.e., ASD vs. HUD) and

(ii) not on the main research question of the present manuscript on potential influences of severe RTAs (i.e., between-subjects variable: accident vs. non-accident group),

(iii) nor on the psychological results reported in the present manuscript,

have been published as a conference proceeding (Schartmüller et al., 2019), which has been presented as work-in-progress at the Auto UI conference 2019.

We want to further mention that another paper has been accepted in the special issue on Future Mobility but

(iv) with a different and *smaller* part of the very large data set primarily focusing on eye tracking which is distinct and not submitted here,

(v) with different research questions,

(vi) ultimately resulting in a different abstract, introduction, results section, discussion, and conclusion.

Conference Proceeding

Schartmüller, C., Weigl, K., Wintersberger, P., Riener, A., & Steinhauser, M. (2019, September). Text Comprehension: Heads-Up vs. Auditory Displays: Implications for a Productive Work Environment in SAE Level 3 Automated Vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 342–354). ACM. https://doi.org/10.1145/3342197.3344547.

Special Issue Paper Schartmüller, C., Weigl, K., Löcken, A., Wintersperger, P., Steinhauser, M., & Riener, A. (2021). Displays for Productive Non-Driving Related Tasks: Visual Behavior and Its Impact in Conditionally Automated Driving. *Multimodal Technol. Interact. 2021, 5*(21), 1-20. doi: https://doi.org/10.3390/mti5040021.

Acknowledgement

We applied the SDC approach for the sequence of authors. This work was conducted under the “Innovative Hochschule” program of the German Federal Ministry of Education and Research (BMBF) under Grant No. 03IHS109A (MenschINBewegung). Participant compensation was paid by the proFOR + program for research projects of the Catholic University of Eichstätt-Ingolstadt. Finally, we are very grateful to two anonymous reviewers for their valuable and insightful comments on a previous version of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aap.2021.106408.

References

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub.

Ansari, S., Akhdar, F., Mandoorah, M., Moutaery, K., 2000. Causes and effects of road traffic accidents in Saudi Arabia. Public Health 114 (1), 37–39. https://doi.org/10.1038/sj.ph.1900610.

Ateş, C., Kaymaz, Ö., Kale, H.E., Tekindal, M.A., 2019. Comparison of test statistics of nonnormal and unbalanced samples for multivariate analysis of variance in terms of type-I error rates. Comput. Math. Methods Med. 1–8. https://doi.org/10.1155/2019/2173638.

Bazilinskyy, P., de Winter, J., 2015. Auditory interfaces in automated driving: an international survey. PeerJ Comput. Sci. 1, e13 https://doi.org/10.7717/peerj-cs.13.

Beggiato, M., Hartwich, F., Krems, J., 2018. Using smartbands, pupillometry and body motion to detect discomfort in automated driving. Front. Hum. Neurosci. 12 https://doi.org/10.3389/fnhum.2018.00338.

Belin, M.A., Johansson, R., Lindberg, J., Tingvall, C., 1997. The vision zero and its consequences. In: Proceedings of the 4th International Conference on Safety and the Environment in the 21 St Century, pp. 23–27.

Benedek, M., Kaernbach, C., 2010. Decomposition of skin conductance data by means of nonnegative deconvolution. Psychophysiology 47 (4), 647–658. https://doi.org/10.1111/j.1469-8986.2009.00972.x.

Benedek, M., Kaernbach, C., 2014. Ledalab Software. http://www.ledalab.de/.

Bener, A., 2005. The neglected epidemic: road traffic accidents in a developing country, State of Qatar. Int. J. Inj. Contr. Saf. Promot. 12 (1), 45–47. https://doi.org/10.1080/1745730051233142225.

Bennun, I.S., Bell, P., 1999. Psychological consequences of road traffic accidents. Med. Sci. Law 39 (2), 167–172. https://doi.org/10.1177/002580249903900212.

Blevins, C.A., Weathers, F.W., Davis, M.T., Witte, T.K., Domino, J.L., 2015. The posttraumatic stress disorder checklist for DSM-5 (PCL-5): development and initial psychometric evaluation. J. Trauma. Stress 28 (6), 489–498. https://doi.org/10.1002/jts.22059.

Brookhuis, K.A., Waard, D.D., Fairclough, S.H., 2003. Criteria for driver impairment. Ergonomics 46 (5), 433–445. https://doi.org/10.1080/001401302/1000039556.

Brown, J.D., Huffman, W.J., 1972. Psychophysiological measures of drivers under actual driving conditions. J. Saf. Res. 4 (4), 172–178.

Chmitorz, A., Kunzler, A., Helmreich, I., Tüscher, O., Kalisch, R., Kubiak, T., Wessa, M., Lieb, K., 2018. Intervention studies to foster resilience – a systematic review and proposal for a resilience framework in future intervention studies. Clin. Psychol. Rev. 59, 78–100. https://doi.org/10.1016/j.cpr.2017.11.002.

Clapp, J.D., Baker, A.S., Litwack, S.D., Sloan, D.M., Beck, J.G., 2014. Properties of the driving behavior survey among individuals with motor vehicle accident-related posttraumatic stress disorder. J. Anxiety Disord. 28 (1), 1–7. https://doi.org/10.1016/j.janxdis.2013.10.008.

Clapp, J.D., Olsen, S.A., Danoff-Burg, S., Hagewood, J.H., Hickling, E.J., Hwang, V.S., Beck, J.G., 2011. Factors contributing to anxious driving behavior: the role of stress history and accident severity. J. Anxiety Disord. 25 (4), 592–598. https://doi.org/10.1016/j.janxdis.2011.01.008.

Clark, H., Feng, J., 2017. Age differences in the takeover of vehicle control and engagement in non-driving-related activities in simulated driving with conditional automation. Accid. Anal. Prev. 106, 468–479. https://doi.org/10.1016/j.aap.2016.08.027.

Cohen, J., 1992. A power primer. Psychol. Bull. 112 (1), 155–159.

Collet, C., Clarion, A., Morel, M., Chapon, A., Petit, C., 2009. Physiological and behavioural changes associated to the management of secondary tasks while driving. Appl. Ergon. 40 (6), 1041–1046. https://doi.org/10.1016/j.apergo.2009.01.007.

Cortina, J.M., 1993. What is coefficient alpha? An examination of theory and applications. J. Appl. Psychol. 78 (1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16 (3), 297–334. https://doi.org/10.1007/bf02310555.

Cronbach, L.J., Rajaratnam, N., Gleser, G.C., 1963. Theory of generalizability: a liberalization of reliability theory. Br. J. Stat. Psychol. 16 (2), 137–163. https://doi.org/10.1111/j.2044-8317.1963.tb00206.x.

Daneman, M., Carpenter, P.A., 1980. Individual differences in working memory and reading. J. Verbal Learn. Verbal Behav. 19 (4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6.

Dixit, V., Xiong, Z., Jian, S., Saxena, N., 2019. Risk of automated driving: Implications on safety acceptability and productivity. Accid. Anal. Prev. 125, 257–266. https://doi.org/10.1016/j.aap.2019.02.005.

Frank, A., 1993. Psychiatric consequences of road traffic accidents. Often disabling and unrecognised. Br. Med. J. 307 (6914), 1283.

Gopalakrishnan, S., 2012. A public health perspective of road traffic accidents. J. Fam. Med. Prim. Care 1 (2), 144–150. https://doi.org/10.4103/2249-4863.104987.

Haghzare, S., Campos, J.L., Bak, K., Mihailidis, A., 2021. Older adults' acceptance of fully automated vehicles: effects of exposure, driving style, age, and driving conditions. Accid. Anal. Prev. 150, 105919 https://doi.org/10.1016/j.aap.2020.105919.

Hart, S.G., 2006. Nasa-Task Load Index (NASA-TLX); 20 years later. Proc. Hum. Factors Ergon. Soc. Ann. Meet. 50 (9), 904–908. https://doi.org/10.1177/154193120605000909.

Herrman, H., Stewart, D.E., Diaz-Granados, N., Berger, E.L., Jackson, B., Yuen, T., 2011. What is resilience? Can. J. Psychiatry 56 (5), 258–265. https://doi.org/10.1177/070674371105600504.

Hobbs, M., Mayou, R., Harrison, B., Worlock, P., 1996. A randomised controlled trial of psychological debriefing for victims of road traffic accidents. BMJ 313 (7070), 1438–1439. https://doi.org/10.1136/bmj.313.7070.1438.

Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6 (2), 65–70.

Corp, I.B.M., 2017. IBM SPSS Statistics for Windows (Version 25) [Computer software]. IBM Corp, Armonk, NY.

Jian, J.-Y., Bisantz, A.M., Drury, C.G., 1998. Towards an empirically determined scale of trust in computerized systems: distinguishing concepts and types of trust. Proc. Hum. Factors Ergon. Soc. Ann. Meet. 42 (5), 501–505. https://doi.org/10.1177/154193129804200512.

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

Johansson, R., 2009. Vision zero – implementing a policy for traffic safety. Saf. Sci. 47 (6), 826–831. https://doi.org/10.1016/j.ssci.2008.10.023.

Knappe, G., Keinath, A., Bengler, K., Meinecke, C., 2007. Driving simulator as an evaluation tool–assessment of the influence of field of view and secondary tasks on lane keeping and steering performance. In: 20th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration, pp. 07–0262.

Koisaari, T., Utriainen, R., Kari, T., Tervo, T., 2020. The most difficult at-fault fatal crashes to avoid with current active safety technology. Accid. Anal. Prev. 135, 105396 https://doi.org/10.1016/j.aap.2019.105396.

Kroenke, K., Spitzer, R.L., 2002. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr. Ann. 32 (9), 509–515. https://doi.org/10.3928/0048-5713-20020901-06.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9, validity of a brief depression severity measure. J. Gen. Intern. Med. 16 (9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Kuch, K., 1993. Psychiatric consequences of road traffic accidents. Phobias linked to chronic pain. Br. Med. J. 307 (6914), 1283.

Kuch, K., Cox, B.J., Evans, R., Shulman, I., 1994. Phobias, panic, and pain in 55 survivors of road vehicle accidents. J. Anxiety Disord. 8 (2), 181–187. https://doi.org/10.1016/0887-6185(94)90015-9.

Kuch, K., Evans, R.J., Watson, P.C., Bubela, C., Cox, B.J., 1991. Road vehicle accidents and phobias in 60 patients with fibromyalgia. J. Anxiety Disord. 5 (3), 273–280. https://doi.org/10.1016/0887-6185(91)90008-H.

Kun, A.L., Boll, S., Schmidt, A., 2016. Shifting gears: user interfaces in the age of autonomous driving. IEEE Pervasive Comput. 15 (1), 32–38. https://doi.org/10.1109/MPRV.2016.14.

Leppert, K., Koch, B., Brähler, E., Strauß, B., 2008. Die Resilienzskala (RS) – Überprüfung der Langform RS-25 und einer Kurzform RS-13. Klin. Diagnostik u. Eval. 1, 226–243.

Li, Y., Tu, Y., Fan, Q., Dong, C., Wang, W., 2018. Influence of cyber-attacks on longitudinal safety of connected and automated vehicles. Accid. Anal. Prev. 121, 148–156. https://doi.org/10.1016/j.aap.2018.09.016.

Limesurvey Project Team, Schmitz, C., 2021. LimeSurvey: An Open Source Survey Tool. LimeSurvey GmbH.

Lin, Q., Li, S., Ma, X., Lu, G., 2020. Understanding take-over performance of high crash risk drivers during conditionally automated driving. Accid. Anal. Prev. 143, 105543 https://doi.org/10.1016/j.aap.2020.105543.

Liu, P., Du, Y., Xu, Z., 2019. Machines versus humans: People's biased responses to traffic accidents involving self-driving vehicles. Accid. Anal. Prev. 125, 232–240. https://doi.org/10.1016/j.aap.2019.02.012.

Löwe, B., Spitzer, R. L., Zipfel, S., Herzog, W., 2002. PHQ-D Gesundheitsfragebogen für Patienten. Manual, Komplettversion und Kurzform, Autorisierte deutsche Version des 'Prime MD Patient Health Questionnaire (PHQ)', 2nd ed. Pfizer. https://www.klinikum.uni-heidelberg.de/fileadmin/Psychosomatische_Klinik/download/PHQ_Manual1.pdf.

Manyara, C.G., 2016. Combating road traffic accidents in Kenya: a challenge for an emerging economy. In: Koster, M.M., Kithinji, M.M., Rotich, J.P. (Eds.), Kenya After 50: Reconfiguring Education, Gender, and Policy. Palgrave Macmillan US, pp. 101–122. https://doi.org/10.1057/9781137574633_7.

Mayou, R., Bryant, B., 2002. Outcome 3 years after a road traffic accident. Psychol. Med. 32 (4), 671–675. https://doi.org/10.1017/S0033291702005470.

Mayou, R., Bryant, B., Duthie, R., 1993. Psychiatric consequences of road traffic accidents. Br. Med. J. 307 (6905), 647–651. https://doi.org/10.1136/bmj.307.6905.647.

Mayou, R., Ehlers, A., Hobbs, M., 2000. Psychological debriefing for road traffic accident victims: three-year follow-up of a randomised controlled trial. Br. J. Psychiatry 176 (6), 589–593. https://doi.org/10.1192/bjp.176.6.589.

Mayou, R., Tyndel, S., Bryant, B., 1997. Long-term outcome of motor vehicle accident injury. Psychosom. Med. 59 (6), 578–584.

McDonald, A.D., Alambeigi, H., Engström, J., Markkula, G., Vogelpohl, T., Dunne, J., Yuma, N., 2019. Toward computational simulations of behavior during automated driving takeovers: a review of the empirical and modeling literatures. Hum. Factors 61 (4), 642–688.

Murray, J., Ehlers, A., Mayou, R., 2002. Dissociation and post-traumatic stress disorder: two prospective studies of road traffic accident survivors. Br. J. Psychiatry 180 (4), 363–368. https://doi.org/10.1192/bjp.180.4.363.

Museru, L.M., Mcharo, C.N., Leshabari, M.T., 2002. Road traffic accidents in Tanzania: a ten year epidemiological appraisal. East Cent. Afr. J. Surg. 7 (1).

Naqvi, N.K., Quddus, M.A., Enoch, M.P., 2020. Do higher fuel prices help reduce road traffic accidents? Accid. Anal. Prev. 135, 105353 https://doi.org/10.1016/j.aap.2019.105353.

Nees, M.A., Sharma, N., Shore, A., 2020. Attributions of accidents to "human error" in news stories: effects on perceived culpability, perceived preventability, and perceived need for punishment. Accid. Anal. Prev. 148, 105792 https://doi.org/10.1016/j.aap.2020.105792.

Nunnally, J.C., 1978. Psychometric Theory, (2nd ed.). McGraw-Hill.

Pauzie, A., 2015. Head up display in automotive: a new reality for the driver. In: Marcus, A. (Ed.), Design, User Experience, and Usability: Interactive Experience Design. Springer International Publishing, pp. 505–516. https://doi.org/10.1007/978-3-319-20889-3_47.

Pfleging, B., Rang, M., Broy, N., 2016. Investigating user needs for non-driving-related activities during automated driving. In: Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia – MUM'16, pp. 91–99. https://doi.org/10.1145/3012709.3012735.

Pillai, K.C.S., 1955. Some new test criteria in multivariate analysis. Ann. Math. Stat. 26 (1), 117–121. www.jstor.org/stable/2236762.

Rendon-Velez, E., van Leeuwen, P.M., Happee, R., Horváth, I., van der Vegte, W.F., de Winter, J.C.F., 2016. The effects of time pressure on driver performance and physiological activity: a driving simulator study. Transp. Res. F: Traffic Psychol. Behav. 41, 150–169. https://doi.org/10.1016/j.trf.2016.06.013.

SAE, 2018. J3016_201806: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems—SAE International. Society of Automobile Engineers, sae.org. https://www.sae.org/standards/content/j3016_201401/.

SAE J2944, 2015. J2944: Operational Definitions of Driving Performance Measures and Statistics – SAE International. https://www.sae.org/standards/content/j2944_201506/.

Schumacher, J., Leppert, K., Gunzelmann, T., Strauß, B., Brähler, E., 2005. Die Resilienzskala—Ein Fragebogen zur Erfassung der psychischen Widerstandsfähigkeit als Personmerkmal. Z. Klin. Psychol. Psychopathol. Psychother. 53 (1), 16–39.

Seber, G.A.F., 1984. Multivariate Observations. John Wiley & Sons, Inc.

Shuttleworth, J., 2019. SAE standards news: J3016 automated-driving graphic update. SAE J. 2.

Singh, S.K., 2017. Road traffic accidents in India: issues and challenges. Transp. Res. Procedia 25, 4708–4719. https://doi.org/10.1016/j.trpro.2017.05.484.

Sisto, A., Vicinanza, F., Campanozzi, L.L., Ricci, G., Tartaglini, D., Tambone, V., 2019. Towards a transversal definition of psychological resilience: a literature review. Medicina 55 (11). https://doi.org/10.3390/medicina55110745.

Southwick, S.M., Bonanno, G.A., Masten, A.S., Panter-Brick, C., Yehuda, R., 2014. Resilience definitions, theory, and challenges: Interdisciplinary perspectives. Eur. J. Psychotraumatol. 5 https://doi.org/10.3402/ejpt.v5.25338.

Stanton, N., Hedge, A., Hendrick, H.W., Brookhuis, K.A., Salas, E., 2004. Handbook of Ergonomics and Human Factors Methods. Taylor & Francis, London, pp. 201–208.

Taber, K.S., 2018. The use of Cronbach's Alpha when developing and reporting research instruments in science education. Res. Sci. Educ. 48 (6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2.

Taylor, J., Paki, D., 2008. Wanna drive? Driving anxiety and fear in a New Zealand community sample. N. Z. J. Psychol. 37 (2), 31–37.

Taylor, J., Deane, F., Podd, J., 2002. Driving-related fear: A review. Clin. Psychol. Rev. 22, 631–645. https://doi.org/10.1016/s0272-7358(01)00114-3.

Taylor, S., Koch, W.J., 1995. Anxiety disorders due to motor vehicle accidents: nature and treatment. Clin. Psychol. Rev. 15 (8), 721–738. https://doi.org/10.1016/0272-7358(95)00043-7.

Tseng, V.-W.-S., Abdullah, S., Costa, J., Choudhury, T., 2018. AlertnessScanner: what do your pupils tell about your alertness. In: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 1–11. https://doi.org/10.1145/3229434.3229456.

Tsimhoni, O., Green, P., Watanabe, H., 2001. Detecting and Reading Text on HUDs: Effects of Driving Workload and Message Location. Annual ITS America Meeting.

Valent, F., Schiava, F., Savonitto, C., Gallo, T., Brusaferro, S., Barbone, F., 2002. Risk factors for fatal road traffic accidents in Udine, Italy. Acci. Anal. Prev. 34 (1), 71–84. https://doi.org/10.1016/S0001-4575(00)00104-4.

Vollmer, M., 2019. HRVTool – an open-source matlab toolbox for analyzing heart rate variability. In: 2019 Computing in Cardiology (CinC), pp. 1–4. https://doi.org/10.23919/CinC49843.2019.9005745.

Vorko-Jović, A., Kern, J., Biloglav, Z., 2006. Risk factors in urban road traffic accidents. J. Saf. Res. 37 (1), 93–98. https://doi.org/10.1016/j.jsr.2005.08.009.

Wagnild, G.M., Young, H.M., 1993. Development and psychometric evaluation of the resilience scale. J. Nurs. Meas. 1 (2), 165–178.

Wang, H.-M., Huang, S.-C., 2012. SDNN/RMSSD as a surrogate for LF/HF: a revised investigation. Model. Simul. Eng. 2012, e931943 https://doi.org/10.1155/2012/931943.

Weathers, F.W., Litz, B.T., Keane, T.M., Palmieri, P.A., Marx, B.P., Schnurr, P.P., 2013. The PTSD Checklist for DSM-5 (PCL-5) [General Information]. https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp.

WHO, 2020a, February 7. Road Traffic Injuries, WHO Fact Sheet on Road Traffic Injuries Providing Key Facts. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

WHO, 2020b, December 9. The Top 10 Causes of Death. Fact Sheet on the 10 Leading Causes of Death. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.

Wickens, C.D., 2002. Multiple resources and performance prediction. Theor. Issues Ergon. Sci. 3 (2), 159–177. https://doi.org/10.1080/14639220210123806.

Wiedemann, K., Naujoks, F., Wörle, J., Kenntner-Mabiala, R., Kaussner, Y., Neukum, A., 2018. Effect of different alcohol levels on take-over performance in conditionally automated driving. Accid. Anal. Prev. 115, 89–97. https://doi.org/10.1016/j.aap.2018.03.001.

Wintersberger, P., Green, P., Riener, A., 2017. Am I Driving or Are You or Are We Both? A Taxonomy for Handover and Handback in Automated Driving. Ninth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Manchester Village, Vermont. Iowa City 333–339. https://doi.org/10.17077/drivingassessment.1655.

Wu, G., Feder, A., Cohen, H., Kim, J.J., Calderon, S., Charney, D.S., Mathé, A.A., 2013. Understanding resilience. Front. Behav. Neurosci. 7 https://doi.org/10.3389/fnbeh.2013.00010.

Yerkes, R.M., Dodson, J.D., 1908. The relation of strength of stimulus to rapidity of habit-formation. J. Comp. Neurol. Psychol. 18 (5), 459–482. https://doi.org/10.1002/cne.920180503.