Accident Analysis and Prevention 121 (2018) 28–42

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

From partial and high automation to manual driving: Relationship between non-driving related tasks, drowsiness and take-over performance

Frederik Naujoks[a], Simon Höfling[a], Christian Purucker[a,*], Kathrin Zeeb[b]

[a] Würzburg Institute for Traffic Sciences GmbH (WIVW), Robert-Bosch-Straße 4, 97209 Veitshöchheim, Germany
[b] Robert Bosch GmbH, CC-AD/EYF3, Robert-Bosch-Allee 1, 74232 Abstatt, Germany

A R T I C L E I N F O

Keywords:
Automated driving
Controllability
Drowsiness
Distraction

A B S T R A C T

Background: Until the level of full vehicle automation is reached, users of vehicle automation systems will be required to take over manual control of the vehicle occasionally and stay fallback-ready to some extent during the drive. Both, drowsiness caused by inactivity and the engagement in distracting non-driving related tasks (NDRTs) such as entertainment or office work have been suggested to impair the driver's ability to safely handle these transitions of control. Thus, it is an open question whether engagement in NDRTs will impair or improve take-over performance.
Method: In a motion-based driving simulator, 64 participants completed an automated drive that lasted either one or two hours using either a partially or highly automated driving system. In the partially automated driving condition, a warning was issued after several seconds when drivers took both hands off the steering wheel, while the highly automated driving system allowed hands-off driving permanently. Drivers were allowed to bring along their smartphones and to use them during the drive. They engaged in a wide variety of NDRTs such as reading or using social media. At the end of the session, drivers had to react to a sudden lead vehicle braking event. In the partial automation condition, there was no take-over request (TOR) to notify the drivers of the braking vehicle, while in the highly automated condition, the situation happened right after the drivers had deactivated the automation in response to a TOR. The lead time of the TOR was set at 8 s. Driver's level of drowsiness, workload (visual, mental and motoric) from carrying out the NDRT and motivational appeal of the NDRT right before the control transition were video-coded and used to predict the outcome of the braking event (i.e., reaction and system deactivation times, minimal Time-to-collision (TTC) and self-reported criticality) with a multiple regression approach.
Results: In the partial automation condition, reaction times to the braking vehicle and situation criticality as measured by the minimum TTC could be well predicted. Main predictors for increased reaction time were drowsiness and motivational appeal of the NDRT. However, visual and mental demand associated with NDRTs did decrease reaction time, suggesting that the NDRT helped the drivers to maintain alertness during the partially automated drive. Accordingly, drowsiness and motivational appeal of the NDRT increased situation criticality, while cognitive load due to the NDRT decreased it. In the highly automated condition, however, it was not possible to predict system deactivation time (in reaction to the TOR), brake reaction time to the braking vehicle and situation criticality by observed drowsiness and NDRT engagement.
Discussion: The results suggest a relationship between the driver's drowsiness and NDRT engagement in partial automation but not in highly automated driving. Several explanations for this finding are discussed. It could be possible that the lead time of 8 s might have given the drivers enough time to complete the driver state transition process from executing NDRTs to manual driving, putting them in a position to be able to cope with the driving event, while this was not possible in the partial automation condition. Methodological issues that might have led to a non-detection of an effect of drowsiness or NDRT engagement in the highly automated driving condition, such as the sample size and sensitivity of the observer ratings, are also discussed.

* Corresponding author.
E-mail addresses: naujoks@wivw.de (F. Naujoks), hoefling@wivw.de (S. Höfling), purucker@wivw.de (C. Purucker), kathrin.zeeb@de.bosch.de (K. Zeeb).

https://doi.org/10.1016/j.aap.2018.08.018
Received 5 February 2018; Received in revised form 1 July 2018; Accepted 1 August 2018
Available online 08 September 2018

1. Introduction

Automated driving relieves the driver from having to continuously control lateral and longitudinal dynamics of the vehicle. However, depending on the level of automation (Gasser et al., 2012), the driver is still required to monitor the driving environment so that she/he can intervene in case of system limits or malfunctions (so-called "partially automated driving"), or to be available as a fallback level when the system reaches its limits (so-called "highly automated driving"). Unlike "fully automated driving", both partially and highly automated driving still require the driver to resume manual driving at least occasionally during the drive.

Several studies have dealt with the question how drivers manage these so-called take-over situations (e.g., Gold et al., 2015, 2013; Körber et al., 2016; Naujoks et al., 2014, 2015; Naujoks et al., 2017a; Zeeb et al., 2016, 2017). Especially the time budget needed to safely deactivate the automation (so-called "take-over time"; Marberger et al., 2017) and to regain control over the vehicle (so-called "control stabilization time"; Marberger et al., 2017) have attracted considerable research interest. In a review of recent studies on take-over times during transitions from highly automated to manual driving, Eriksson and Stanton (2017) concluded that the reported times had a large variety of between 1 and 15 s.

One explanation why it takes drivers some time before they are ready to take over vehicle control may be that drivers need to switch from executing NDRTs to manual driving (Naujoks et al., 2018). Indeed, several studies (de Winter et al., 2014; Jamson et al., 2013; Large et al., 2017; Naujoks et al., 2016; Naujoks and Totzke, 2014) and surveys (König and Neumayr, 2017; Pfleging et al., 2016) have shown that drivers will likely engage in NDRTs while driving in automated mode, such as reading magazines or using smartphones. To explain the time needed to disengage from NDRTs and to re-engage in the driving task, it has been suggested that time-consuming re-configuration processes of the drivers' *sensory* (e.g., taking one's eyes off the NDRT and attending to relevant HMI displays), *motoric* (e.g., freeing one's hands and taking them back on the steering wheel) and *cognitive* state (e.g., re-configuration of mental task sets or response rules) have to be performed by the driver to meet the demands of manually controlling the vehicle (Marberger et al., 2017). Indeed, research from cognitive psychology has repeatedly demonstrated that switching tasks goes along with costs in the form of increased reaction times or error rates (Altmann and Trafton, 2004; Kiesel et al., 2010; Trafton et al., 2003).

Empirical research indicates that switching costs affect taking over vehicle control after automated driving as well. Take-over performance when switching from automated driving while working on a NDRT to manual driving has been investigated in several studies, usually showing increased take-over times as compared to reference drives without NDRT (Dogan et al., 2017; Körber et al., 2016; Merat et al., 2012; Naujoks et al., 2017a). However, only very few studies have directly compared the impact of different NDRTs against each other. Zeeb et al. (2016) found that drivers who watched a video on a fixed in-vehicle display were slower to respond to take-over requests when compared to a condition without NDRT, but this was not the case when they had to write an email or read a news text. Both, watching a video and reading a news article impaired drivers' lane keeping. In another study (Zeeb et al., 2017), the authors report that motoric task load while reading texts (i.e., taking a tablet into one's hands vs. using a fixed in-vehicle display) increased take-over time and decreased lane keeping directly after control had been taken over. Similar results have been reported by Gold et al. (2015).

While these studies indicate that an increased workload due to the engagement in NDRTs results in increased take-over times and impairments of take-over quality, a state of cognitive underload may equally affect the driver's take-over performance. As a result of the disengagement from driving related activities, drivers might easily get tired when their only task is to monitor the automation. In line with other researchers (e.g., Desmond et al., 1998; Schömig et al., 2015), Vogelpohl and Vollrath (2017) showed in a simulator study that due to the reduced workload drivers experienced fatigue earlier during a highly automated drive than during a manual drive. Consequently, the finding of Miller et al. (2015) that drivers who engage in NDRTs while driving in automated mode show fewer signs of fatigue is not surprising because they might counteract cognitive underload through NDRT engagement. However, in this respect the results in scientific studies are not clear. For example, Neubauer et al. (2014) could show that engagement in NDRTs reduces the perceived stress of boredom in automated driving, but reaction times to critical driving events were not improved, suggesting that engagement in NDRTs does not enhance alertness lastingly.

Taken together, the drivers' state during the automated drive seems to have a great influence on the time needed to re-engage in the driving task. However, the "driver state transition process" (Marberger et al., 2017) is not yet completely understood to date. On the one hand, non-activity could lead to mental underload and drowsiness, resulting in prolonged reaction times and impaired manual driving performance (Jarosch et al., 2017). On the other hand, engagement in NDRTs can lead to an increase in visual, cognitive and motoric load, prolonging the time needed to re-engage in the driving task (Naujoks et al., 2018). It is also possible that a high incentive to continue the NDRT causes drivers to unnecessarily delay their reactions to potentially critical driving events or HMI messages (Ko and Ji, 2018; Wickens et al., 2015). The aim of the present study was to advance existing knowledge on the influence of NDRT engagement and drowsiness on the driver's ability to regain manual control over a highly or partially automated vehicle. The ultimate aim of the study was to gain insights into how driver monitoring systems, such as inferring driver state information from gaze behavior, body postures and usage of smartphones (Braunagel et al., 2017a, 2017b; Louw et al., 2016; Trivedi et al., 2007) could be used to predict driver's reactions to imminent driving situations when switching from automated to manual control.

2. Method

In this simulator study, drivers ultimately encountered an imminent lead vehicle braking scenario they had to react to when resuming control either from a partially or highly automated vehicle. During the preceding automated drive, the participants were free to use their brought-along smartphones and to engage in whatever activity they felt safe enough to do. The drives either lasted one or two hours until the target situation was reached. The aim of the study was to assess the relationship between the drivers' reactions to the braking lead vehicle and their previous NDRT engagement and drowsiness level.

2.1. Experimental design and sample

The study was carried out in a 2 × 2 between-subjects design with $n = 16$ participants per cell ($N = 64$; independent variables: "level of automation" (partially vs. highly automated), and "duration" (1 h vs. 2 h), see Table 1). Drivers were randomly assigned to the experimental groups. The participants were recruited from the test driver panel of the Würzburg Institute for Traffic Sciences (WIVW GmbH) and had participated in an extensive simulator training program prior to the study.

Table 1
Experimental setup of the simulator study.

N = 64		Level of automation	
		Partially Automated	Highly Automated
Duration	1 h	n = 16	n = 16
	2 h	n = 16	n = 16

Table 2
Age distribution by level of automation and duration of automated drive.

Condition	Duration	*Min*	*Max*	*M*	*SD*
Partially Automated	1 h	22	54	35.75	11.27
	2 h	22	56	30.81	9.57
Highly Automated	1 h	24	60	37.44	12.20
	2 h	22	50	31.00	8.43

The inclusion criterion was that participants possessed a smartphone, which they should bring along, fully charged to last for the experimental session. 34 participants were male and 30 female, the mean age was 33 years ($SD = 10$). Parameters of the age distribution per experimental condition are shown in Table 2.

2.2. Test environment

The study was conducted using the moving-base driving simulator at the Würzburg Institute for Traffic Sciences (WIVW GmbH, see Fig. 1). The driving simulation software SILAB developed at the WIVW was used for environment visualization as well as for simulation of assistance systems, traffic and vehicle dynamics. The integrated vehicle's console contains all the necessary instrumentation and is identical with a production type BMW 520i with automatic transmission. In order to simulate a realistic steering torque, a servo motor based on a steering model is used. The motion system uses six degrees of freedom and can briefly display a linear acceleration up to 5 m/s^2 or 100°/s^2 on a rotary scale. It consists of 6 electro-pneumatic actuators (stroke + − 60 cm; inclination +/− 10°). Three LCD projectors are installed in the dome of the simulator and provide the projection. Three channels provide a 300° screen image. LCD displays serve as exterior and interior mirrors.

2.3. Automated driving systems

2.3.1. Partially automated system

The partial automation was labeled “Highway Assistant”. The human-machine interface (HMI) was positioned in the upper center console and resembled HMIs found in today’s production vehicles that are equipped with ACC and steering assistance functions. As shown in Fig. 2, the visual HMI informed the driver when the partially automated system could be activated (“Highway Assistant available”). Pushing a button on the steering wheel activated the partial automation (HMI: “Highway Assistant active”) and it took over longitudinal and lateral vehicle guidance, keeping it at a set speed of 130 km/h. The drivers could thus take their feet off the pedals. In order to make lane keeping appear more realistic (as would be expected in real-world systems), the simulated lateral guidance was not perfect and had a small lateral offset of a few decimeters around the lane center. This was done to simulate a level of vehicle automation that supports the driver in the dynamic driving task, but does not (unlike higher levels of vehicle automation) perform it completely without the need for human supervision (SAEJ3016). The system could be deactivated by pressing the activation button again, or by braking, or by steering. Accelerating led to overriding the set speed. A visual-auditory hands-off warning was displayed when no hands-on signal was detected for a certain time that varied randomly between 15 s and 30 s. The hands-on signal was recorded by pressure sensors in the steering wheel. When the hands-off warning was displayed, the driver had to put the hands on the steering wheel within a timeframe of 10 s; otherwise the partial automation started decelerating and displayed a take-over request (TOR). The remaining time to take the hands back to the steering wheel was visualized by a blue timer in the middle of the HMI. In case of a sensor failure the system was immediately deactivated and the driver was informed by a visual indication (“Please take over! Highway assistant is deactivated”) and an acoustic warning signal.

2.3.2. Highly automated system

The highly automated driving system was labeled “Highway pilot”. As in the partial automation condition, a visual HMI in the upper center console (“Highway Pilot available”) informed drivers that the highly automated system could be activated. The activation and deactivation of the highly automated system was just like in partially automated driving. The activation was also indicated to the drivers (“Highway Pilot active”). Contrary to the partial automation, the system executed lateral guidance perfectly (i.e., there was no lateral offset) and drivers were not required to leave their hands on the steering wheel. In case of system limits or failures, a visualization accompanied by an acoustic warning signal prompted the driver to regain control within a timeframe of 8 s (“Please take over! Highway Pilot is deactivated”). During the take-over mode, the automation stayed active until the driver deactivated it. The remaining time was also visualized by the blue timer in the middle of the HMI.

2.4. Procedure

Having been welcomed, participants signed a confidentiality agreement and gave informed consent. Afterwards, the experimenter explained, depending on the experimental condition, that the purpose of the experiment was to get to know and evaluate a novel partially or highly automated system. Furthermore, participants were given the institute’s public WIFI key. This should guarantee that participants could use internet applications with their smartphones. After being seated in the simulator, participants were given information about the function and system limits of the automated systems. In the partially automated group, the experimenter emphasized that the system needed to be permanently monitored. Conversely, in the highly automated group the experimenter emphasized that participants could fully engage in NDRTs without any need to monitor, because the system would give a timely TOR if necessary. Moreover, participants were instructed how to activate and deactivate the system. In a test drive that took about ten minutes, participants were familiarized with the systems and they could practice the handling (i.e., activation and deactivation). If necessary, the experimenter, who was seated in a separate operator room, supported participants by giving advices using the microphone.

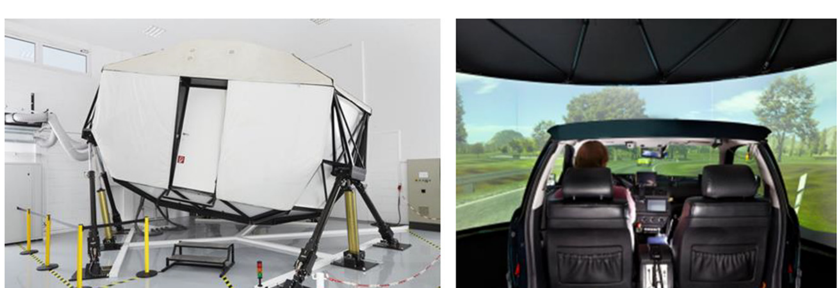

Fig. 1. The WIVW moving-based driving simulator. Hexapod movement system (left) and simulator interior with vehicle mock-up and video projection (right).

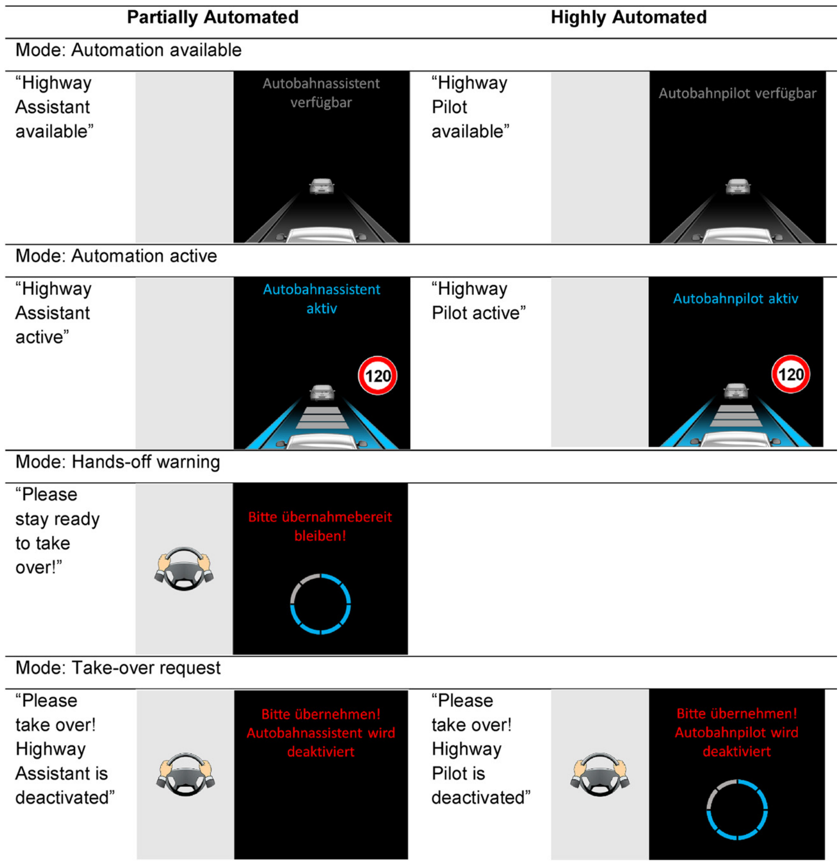

Fig. 2. HMI of the partially (left) and the highly (right) automated systems, displaying various important system states.

After that, the experimenter stopped the simulation to give further instructions. Whereas participants in the partially automated driving condition were told that brought along smartphones were only allowed when using a pre-installed cell phone holder in the center console, participants in the highly automated driving condition were told that they could choose either to use the fixture or to keep the device in their hands. Besides, participants in the highly automated condition were explicitly told that they could also use magazines that were previously placed on the passenger seat. In the partially automated group the magazines were also placed on the passenger seat, but were not mentioned by the experimenter. Finally, participants were asked to stay on the right lane and the main experiment started, if they had no further questions. During the main experiment, no interaction with the participants via the microphone took place until the test situation at the end of the drive was completed.

2.5. Test situation

The test course of the main experiment was set on a three-lane highway with only little traffic. The traffic conditions (e.g., traffic density, velocity, etc.) were kept constant across the participants.

After a 10 min manual drive, the respective automation could be activated. Over the course of the next 50 min or 110 min, the drivers were required to take back manual vehicle control several times, depending on the experimental condition. In the one-hour condition, the drivers in the partially automated drive were required to take back manual vehicle control three times, but only one time in the highly automated condition:

- A construction site made the partial automation system fail and required manual intervention. The highly automated system, however, managed this situation on its own.
- A situation with missing lane markings triggered a TOR in the partially automated drive, but not in the highly automated driving.
- A sudden sensor failure on a straight road segment required both drivers in partial and highly automated condition to take-over manual vehicle control.

In the two-hour condition, the construction site and the missing lanes event appeared twice. Furthermore, two further sensor failures occurred which required the driver to take over both in partial and highly automated driving. The events were presented in randomized order. In the partially automated condition, drivers had to intervene approximately every 15 min. In the highly automated drive, they had to take back manual vehicle control every 30 min. The number of events requiring a driver intervention was different in the partial and high automation condition to reflect different performance levels of both systems.

After 60 or 120 min, respectively, the ***target event*** took place. As the driving was mostly performed by the vehicle automation, there was only little variation of the study duration before the target event between the participants in the two experimental groups. A vehicle approached the participant's car on the left lane, overtook and kept on driving on the middle lane until it was 15.75 m ahead of the ego vehicle and had the same velocity of 130 km/h. After a trigger signal, the vehicle performed a cut-in and braked with a deceleration of 4 m/s^2 (see Fig. 3). In case of a driver non-reaction, it came to a collision after 3.73 s. The trigger signal varied between the two experimental conditions:

- Partially automated condition: As soon as the vehicle reached its position on the left lane, taking at least one hand to the steering

Fig. 3. Multiplexed camera view of the target event (cut-in event).

How fatigued is the driver?				
Not fatigued	Slightly fatigued	Moderately fatigued	Very fatigued	Extremely fatigued
0	1	2	3	4
- Alert behavior - Normal facial tone - Fast eye blinks - Short ordinary glances - Occasional body movements and gestures	- Slightly longer glances - Eye blinks a little slower	- Mannerisms like rubbing face or eyes, scratching, facial contortions, moving restlessly - Alternatively: slower closures, decreased facial tone, glassy-eyed, long glances	- Eyelid closures of 2 to 3 s or longer - improper focusing of eyes, cross-eyed look - low facial tone - Lack of apparent activity or large isolated movements	- Eyelid closures of 4 s and more - intervals of dozing

Fig. 4. Fatigue scale based on Wierwille and Ellsworth (1994) and Wiegand et al. (2009).

wheel or deactivating the automation triggered the cut-in. The automation did not react to the cutting-in vehicle, simulating a sensor failure.

- Highly automated condition: As soon as the vehicle reached its position on the left lane, a TOR was presented, giving the driver 8 s to take back manual control over the simulated vehicle. Any deactivation of the automation triggered the cut-in event.

The test situation required a fast reaction in order to prevent a collision and was rather challenging for the driver. Also no additional collision avoidance systems were available. This was done on purpose in order to analyze the drivers' behavior in a worst-case scenario.

This trigger procedure ensured that participants in both, the partially automated condition as well as the highly automated condition experienced the very same manual and situational demand. It should be nonetheless emphasized, that while the drivers in the highly automated condition had explicitly deactivated the system in the situation (and therefore experienced the situation in manual driving mode), this was not the case for the drivers in the partially automated condition (who experienced a system boundary).

2.6. Measurement of drowsiness and NDRT engagement

Two raters independently evaluated the drivers' drowsiness level and their engagement in the NDRT. Video recordings were multiplexed and showed frontal, rear and legroom perspectives of the participants (Fig. 3). Following Wierwille and Ellsworth (1994), raters evaluated the last 15 s before the onset of the target event. The rating dimensions were:

- Fatigue was rated using the five-point observer rating of Wierwille and Ellsworth (1994) together with the behavioral indicators proposed by Wiegand, McClafferty, McDonald and Hanowski (2009); the scale is shown in Fig. 4.
- Workload induced by the NDRT was rated independently for the rating dimensions visual, mental and motoric workload on a five-point scale shown in Fig. 5.
- Motivation to engage in NDRTs was also rated on a five-point scale depicted in Fig. 6.

The newly developed scales included verbal labels of the respective rating dimension and anchors that were derived from theoretical considerations. Apart from the anchors on the rating scale (i.e., "no NDRT", "looking at a map" and "playing a smartphone game"), the raters had no further pre-classification at hand. To enhance inter-rater reliability, an iterative training consisting out of two parts was conducted prior to the rating of the target events: First of all, the scales were explained and videos were shown to illustrate behavioral indicators of fatigue. Afterwards, the raters evaluated 15 s before the onset of the construction site event. Non-agreements between the raters were discussed.

2.7. Dependent measures of take-over performance

2.7.1. Take-over reaction time

In partially automated driving, braking reaction time in the target event was recorded as a measure of take-over time. This reaction time was defined as the difference between the moment the vehicle started to change lanes and the point when the brake pedal was applied. As drivers in highly automated driving already executed the driving task manually up to the time of the cut-in, here the system deactivation time was recorded as a measure of take-over time (Marberger et al., 2017). Driver's brake reaction times were additionally analyzed as a measure of take-over performance.

How high is the visual workload on the driver?				
Not at all	Little	Moderate	Very	Extreme
0	1	2	3	4
e.g., permanent monitoring of the road		*e.g., using a navigation system*		*e.g., permanent watching of a video*

How high is the mental workload on the driver?				
Not at all	Little	Moderate	Very	Extreme
0	1	2	3	4
e.g., pushing the activation button of the system		*e.g., following the directions of a navigation system*		*e.g., playing a smartphone game*

How high is the motoric workload on the driver?				
Not at all	Little	Moderate	Very	Extreme
0	1	2	3	4
e.g., no NDRT, hands on steering wheel		*e.g., searching for radio channels*		*e.g., searching for an item on rear seat*

Fig. 5. Visual, mental and motoric workload scales. Theory-based examples of NDRTs that induce different levels of workload are used to anchor very low, average and very high workload levels.

2.7.2. Minimum time-to-Collision (TTC)

As an objective measure of take-over quality the minimum time-to-collision was chosen. The TTC, which indicates the time left until a collision occurs if both vehicles continue on converging trajectories at their current speed, is calculated at each measuring point. Minimum TTC can be seen as a quality measure of take-over performance, because low values indicate either a slow braking reaction time or an inadequately low braking force (Radlmayr et al., 2014). Thus, small minimum TTCs indicate critical situations. The formula which was used to compute TTC is shown in the following (Hayward, 1972, d = distance to vehicle in front, v_{rel} = difference in velocity between vehicle in front and following vehicle, note that TTC is only calculated when the following vehicle has a higher speed than the vehicle in front):

$$TTC = \frac{d}{V_{rel}}$$

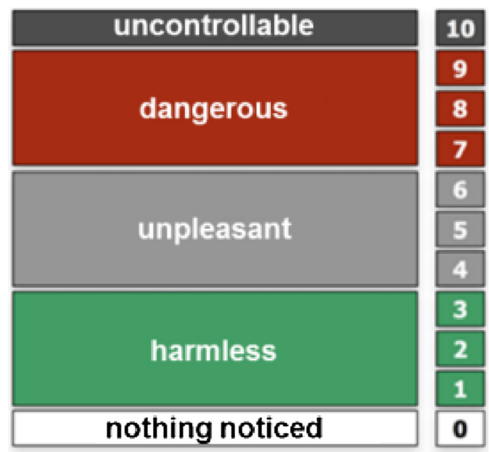

Fig. 7. Subjective Rating of Criticality as proposed by Neukum and Krüger (2003).

2.7.3. Subjective rating of criticality

The experienced criticality was measured using subjective ratings of the participants on the "scale of criticality assessment of driving and traffic situations" (Neukum and Kröger, 2003, see Fig. 7). This one-dimensional scale distinguishes between the five categories "uncontrollable", "dangerous", "unpleasant","harmless" and "nothing noticed" and further allows participants to express tendencies towards other grades by subdividing the middle categories into three levels. Each driver was asked to assess the criticality of the situations on this scale.

3. Results

3.1. Observer ratings prior to target event

3.1.1. Rater reliability analysis

As a first step, the reliability of the observer ratings of drowsiness and NDRT engagement were assessed. As each event was rated by all raters and there was only a single measurement per rating, an intra-class correlation coefficient (model 3, single measurement; ICC 3,1) was used in accordance with Shrout and Fleiss (1979). Contrary to Pearson correlations, the ICC (that is closely related to weighted kappa, Fleiss and Cohen, 1973) does not only provide information about the strength of the relation between two raters; it can also be used to assess the degree to which a rating scale is able to differentiate between participants with diverging scores, taking into account the degree to which two or more raters reach similar conclusions. This is to prevent the shortcoming of Pearson correlations that perfect linear correlations can be achieved even though not one single absolute agreement exists, if one rater systematically differs from the other rater.

As shown in Table 3, the ICC values of all rating dimensions indicated a satisfying degree of reliability. However, according to Koo and Li (2016) confidence intervals need to be considered additionally to interpret inter-rater reliability. These measures indicate a good to excellent reliability for visual workload, a moderate to good reliability for mental workload and motivation, a poor to good reliability for motor workload and a poor to moderate reliability for fatigue.

To understand why the confidence interval for fatigue and motoric workload exhibits such a wide range, a closer look was taken at the degree to which rating scores between raters were identical or differed. It appeared that 96.9% of the ratings on fatigue, 93.8% of the rating on visual workload, 98.4% of the ratings on mental workload, 92.2% of motor workload, and 95.4% of motivation ratings were identical or

How high is the motivational incentive of the NDRT?				
Not at all	Little	Moderate	Very	Extreme
0	1	2	3	4
e.g., no NDRT		*e.g., looking at a map*		*e.g., playing a smartphone game*

Fig. 6. Motivation scale. Theory-based examples of NDRTs with different motivational incentives are used to anchor very low, average and very high motivation levels.

Table 3
Inter-rater reliability (ICC values) for the different rating dimensions.

Rating	r_{ICC}	95% Confidence interval Lower Limit	Upper limit	p-value
Fatigue	.60	.41	.73	< .001
Visual workload	.88	.78	.93	< .001
Mental workload	.81	.70	.88	< .001
Motor workload	.61	.26	.79	< .001
Motivation	.77	.64	.86	< .001

Table 4
Overview of the number of rating discrepancies. Percentages are indicated in brackets.

	Difference between raters			
Rating	0	1	2	3
Fatigue	36 (56.3%)	26 (40.6%)	1 (1.6%)	1 (1.6%)
Visual workload	38 (59.4%)	22 (34.4%)	3 (6.3%)	0 (0.0%)
Mental workload	40 (63.5%)	22 (34.9%)	1 (1.6%)	0 (0.0%)
Motor workload	23 (35.9%)	36 (56.3%)	5 (7.8%)	0 (0.0%)
Motivation	28 (43.8%)	33 (51.6%)	3 (4.7%)	0 (0.0%)

were within one score of each other (Table 4). These results indicate that the unsatisfying measures of statistical uncertainty are caused by single cases that were interpreted differently and not by a constant disagreement between raters. However, it should still be emphasized that these outliers have a substantial impact on the lower bounds of the confidence limits reported in Table 3.

3.1.2. Fatigue

For the following analysis, the observer ratings were averaged across both raters. A full-factorial ANOVA with the independent variables (IVs) “level of automation” and “duration” and the dependent variable (DV) “fatigue ratings” was conducted. Degrees of freedom (*df*), F-value (*F*) and p-value (*p*) are reported, as well partial eta-squared (η^2) as a measure of effect size.

On average, drivers in partially and highly automated driving exhibited very similar patterns of fatigue during the last 15 s before the event onset. As can be seen in Fig. 8, most of the drivers in partially and highly automated driving got only slightly or moderately fatigued. Only one driver per condition was rated as being extremely fatigued. Neither the duration of the automated drive nor the interaction of the level of automation and the duration impacted fatigue (Table 5).

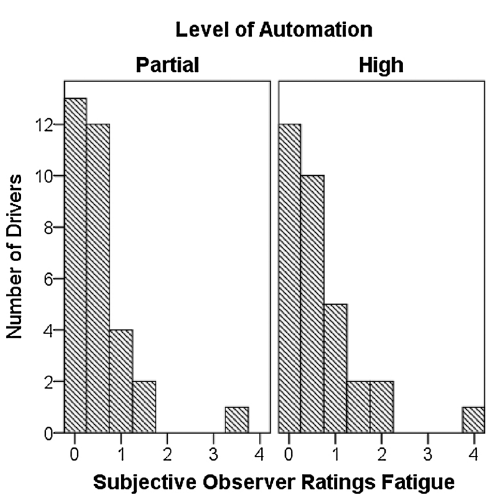

Fig. 8. Distribution of fatigue ratings in partially and highly automated driving in the 15 s before the target event onset.

Table 5
ANOVA results; DV: fatigue ratings; IVs: level of automation and duration.

	df	*F*	*p*	η^2
(A) Level of Automation	1	0.54	.466	.009
(B) Duration	1	1.92	.171	.031
A x B (interaction)	1	1.92	.171	.031
Error	60			

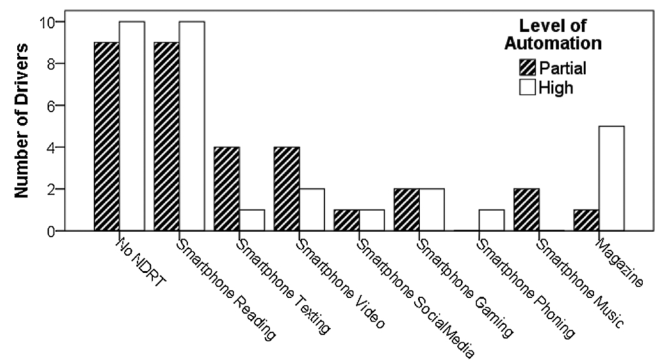

Fig. 9. Overview of NDRTs in partially and highly automated driving in the 15 s before the target event onset.

3.1.3. Observed engagement in NDRTs, workload and motivation ratings

As depicted in Fig. 9, during the last 15 s before the onset of the target event, 9 (28.13%) drivers in partially automated driving and 10 drivers (31.25%) in highly automated driving did not engage in any NDRT. One driver (3.13%) in partially automated driving and 5 drivers (15.63%) in highly automated driving read a magazine. Most of the drivers used their smartphone (70.74% in partially automated driving, 53.12% in highly automated driving). In doing so, the majority read (e.g., surfing on the Internet, reading emails, etc.), while others wrote messages, watched videos, used social media, or played games. Only one driver used his smartphone to make phone calls.

According to the observers (see Fig. 10), visual workload resulting from the engagement in NDRTs was moderate and mental workload was low in both levels of automation. Drivers in highly automated driving exhibited a moderate motoric workload, whereas motoric workload was only low in partially automated driving; the effect of the automation was found to be statistically significant (see Table 6).

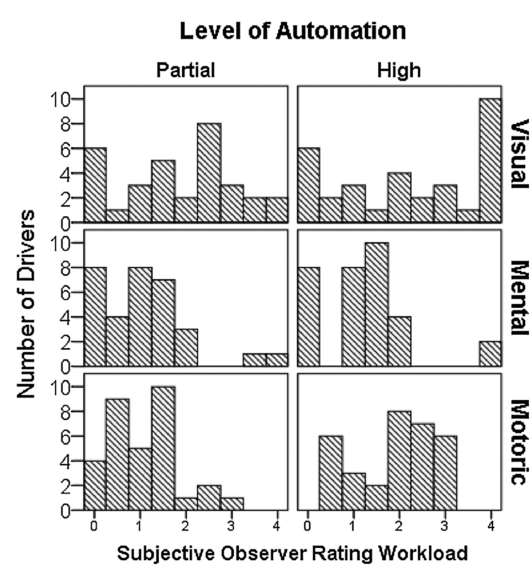

Fig. 10. Distribution of visual, mental and motoric workload ratings in partially and highly automated driving in the 15 s before the target event onset.

Table 6
ANOVA results; DV: workload ratings; IVs: level of automation and duration.

	df	*F*	*p*	η^2
Observed visual workload				
(A) Level of Automation	1	1.12	.294	.018
(B) Duration	1	1.12	.294	.018
A x B (interaction)	1	0.12	.726	.002
Error	60			
Observer mental workload				
(A) Level of Automation	1	0.39	.533	.007
(B) Duration	1	0.25	.618	.004
A x B (interaction)	1	0.06	.803	.001
Error	60			
Observer motoric workload				
(A) Level of Automation	1	15.19	< .001	.202
(B) Duration	1	0.20	.655	.003
A x B (interaction)	1	0.81	.372	.013
Error	60			

Despite these general observations, there was a considerable variance on all workload dimensions. The duration of the automated drive did not influence workload ratings, nor did the interaction between level of automation and duration (see Table 6).

Participants in partially and highly automated driving did not differ in terms of observer ratings of motivation to carry out the NDRT. On average motivation was low to moderate in both levels of automation. Again, it became evident that there was a considerable variance within the sample (Fig. 11). There were neither significant main effects of the level of automation or the automated driving time found on the motivation ratings, nor an interaction of both factors (Table 7).

3.2. Measures of take-over performance

As a next step, the driver's take-over performance in the target event is analyzed per experimental condition, comparing the effects of automated drive duration using t-tests for independent samples. T-value (t), degrees of freedom (df), p-value (p) as well as cohen's d as a measure of effect size are reported. The highly and partially automated drives are not directly compared against each other for two reasons. First, while in partially automated driving the take-over reaction time was determined by the braking reaction time, in highly automated driving it was determined by the system deactivation time (note that in the highly automated driving condition, the driver braking reaction was required after the system had been previously deactivated). Second, take-over situations were not directly comparable because drivers in partially automated driving needed to recognize the need for manual intervention themselves, while in highly automated driving, a TOR was presented prior to the event.

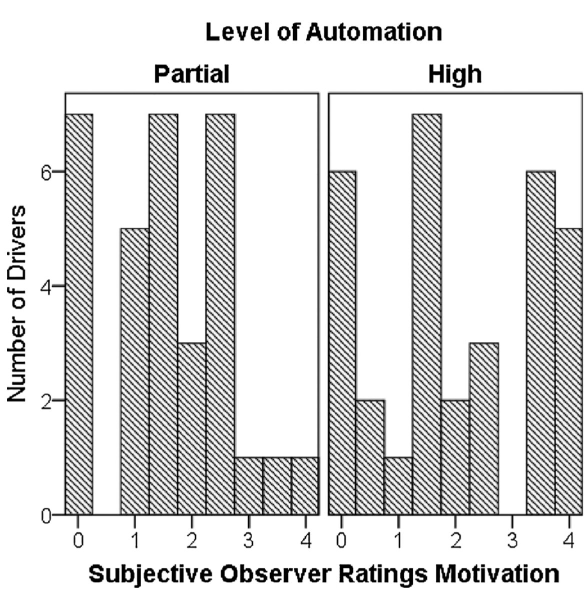

Fig. 11. Distribution of motivation ratings in partially and highly automated driving in the 15 s before the target event onset.

Table 7
ANOVA results; DV: motivation ratings; IVs: level of automation and duration.

	df	*F*	*p*	η^2
(A) Level of Automation	1	2.24	.140	.036
(B) Duration	1	0.52	.472	.009
A x B (interaction)	1	0.67	.415	.011
Error	60			

In partially automated driving $n = 1$ participant failed to react cutting-in vehicle. This particular driver was looking at the roadway right before the event took place, but did not react even after a collision occurred. Since this is not a realistic driver reaction it was assumed that this driver was not sufficiently immersed in the driving simulation. Therefore, this driver was excluded from the analysis of take-over reaction times. On average, participants in partially automated driving started braking after 1.86 s ($SD = 0.64$). The duration of the automated drive did not influence braking reaction time ($t(29) = -0.16$, $p = .875$, $d = .058$). 11 participants in the partially automated driving condition initiated a lane change while deactivating the system with braking; the remaining 20 participants performed solely a braking maneuver.

On average, participants in highly automated driving deactivated the system 3.01 s ($SD = 1.44$) after the TOR was issued. The deactivation triggered the onset of the target event (i.e., the cutting-in vehicle). Averagely, the participants started braking 1.04 s ($SD = 0.67$) after the onset of the target event. Again, the duration of the automated drive did not influence take-over reaction times ($t(30) = 1.24$, $p = .225$, $d = .438$) and brake reaction times ($t(30) = 0.79$, $p = .436$, $d = .278$, see Fig. 12). Whereas the majority of 27 participants performed solely a braking maneuver, a fraction of 5 participants in the highly automated driving condition initiated a lane change as well.

There were $n = 7$ collisions with the front vehicle in partially automated driving. Mean minimum TTC in partially automated driving was 2.02 s ($SD = 1.33$, see Fig. 13). In highly automated driving there was no collision and the drivers had a mean minimum TTC of 4.20 s ($SD = 2.03$). The duration did neither influence minimum TTC in partially automated driving ($t(30) = 0.49$, $p = .625$, $d = .173$), nor in highly automated driving ($t(30) = -0.28$, $p = .785$, $d = .099$). Distributions of the minimum TTC in partially automated driving and highly automated driving are displayed in Fig. 14.

Participants in partially automated driving rated the situation as "dangerous" ($M = 8.56$, $SD = 1.48$), while participants in highly automated driving rated it between "unpleasant" and "dangerous" ($M = 6.56$, $SD = 2.03$, see Fig. 15). The duration of the automated drive did not influence criticality ratings neither in partially automated driving ($t(30) = -0.95$, $p = .348$, $d = .336$), nor in highly automated driving ($t(30) = 0.69$, $p = .495$, $d = 0.244$).

3.3. Prediction of take-over performance

Multiple regression analyses were conducted to predict brake reaction times to the braking vehicle, system deactivation time in highly automated driving, minimum TTC and subjective criticality ratings. The predictors included were the averaged observer ratings of the two raters on fatigue, visual workload, mental workload, motoric workload and motivation to accomplish NDRTs. These predictor variables were included simultaneously into the regression model. As take-over measures were not comparable between automation levels, these analyses were

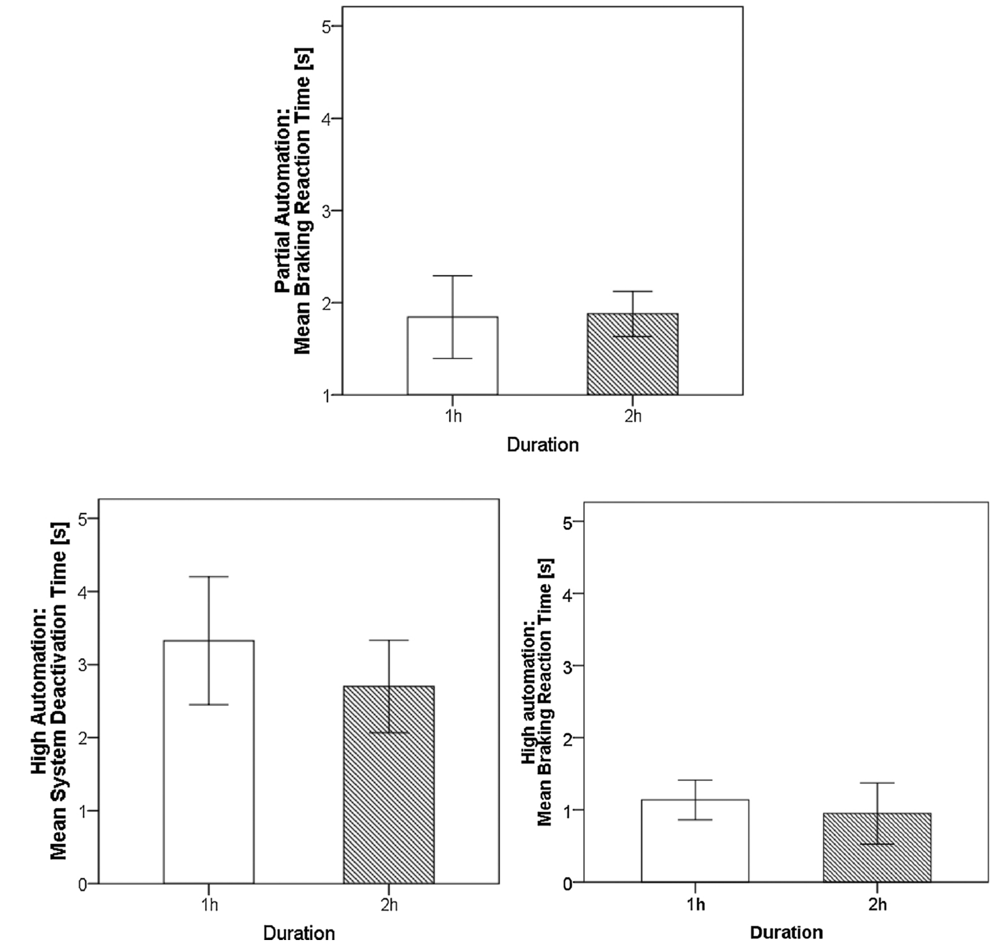

Fig. 12. Comparison of take-over reaction times and brake reaction times in the target event between the 1h- and 2h-duration for partially and highly automated driving. The error bars represent the 95%-confidence intervals.

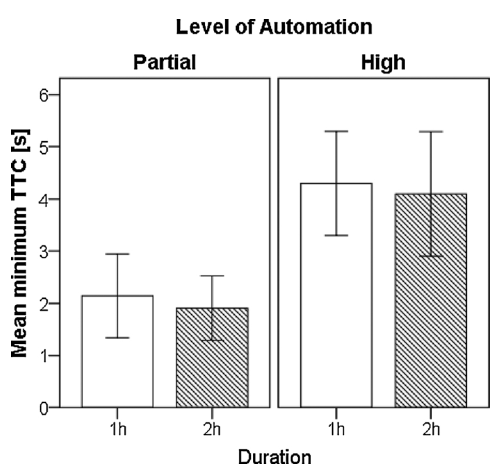

Fig. 13. Comparison of minimum TTC in the target event between the 1h- and 2h-duration for partially and highly automated driving. The error bars represent the 95%-confidence intervals.

conducted separately for partially and highly automated driving. Assumptions of the multiple regression were checked prior to the analysis (see Appendix A).

We report the unstandardized regression coefficient (*B*) as well as its standard error (*SE B*), and 95% confidence limits (*CI*). The standardized regression coefficient (β) is also reported. Model fit is evaluated by reporting the *F*-value, multiple and adjusted R^2.

3.3.1. Predicting take-over reaction time

3.3.1.1. Partially automated driving (Braking reaction time). As summarized in Table 8, the regression model calculated to predict braking reaction time in partially automated driving yielded a significant regression equation. As expected, both fatigue ($t(25) = 3.76$, $p = .001$) and motivation ($t(25) = 5.17$, $p < .001$) contributed significantly to the prediction of brake reaction time to the braking vehicle. An increase of one increment in fatigue ratings slowed the braking reaction time by 0.53 s, respectively 0.99 s in motivation ratings.

Moreover, visual ($t(25) = -2.85$, $p = .009$) and mental ($t(25) = -2.29$, $p = .031$) workload contributed significantly to the model. Interestingly, an increase of one increment in visual workload ratings accelerated braking reaction time by 0.45 s, respectively 0.29 s in mental workload rating. This somewhat surprising finding may be due to the fact that the observed visual and mental workload levels usually not surpassed “moderate” workload, which may be just enough to keep the driver on an acceptable arousal level during the automated drive that could have speeded up the reaction times.

The motoric workload rating did not significantly contribute to predicting brake reaction times, ($t(25) = -0.56$, $p = .582$), which could be explained by the fact that the participants were not allowed to take the smartphone in their hands, avoiding excessive motoric workload. The partial regression plots (showing the unique effects of the predictors on the response variable) of the significant predictors are illustrated in Fig. 16.

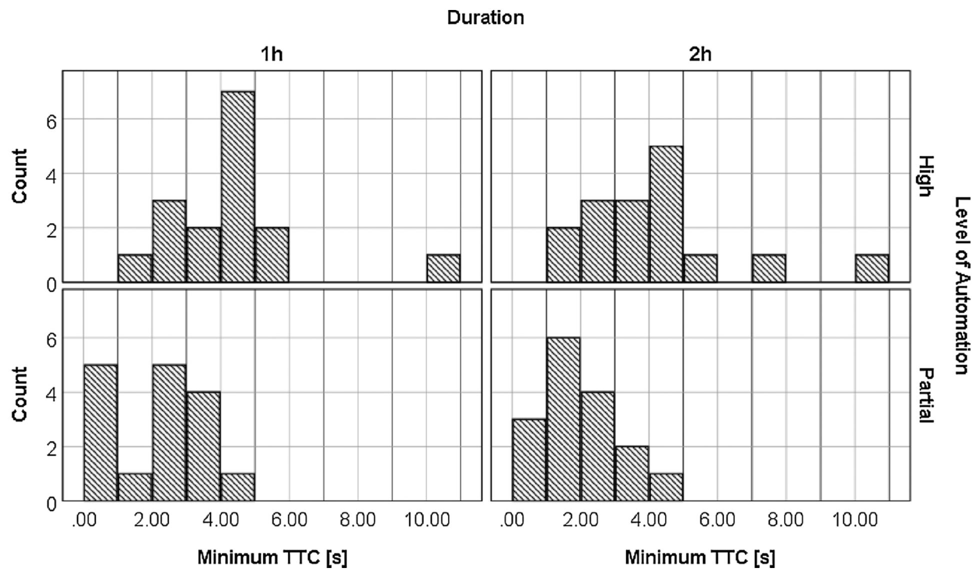

Fig. 14. Count distribution of minimum TTC in the target event between the 1h- and 2h-duration for partially and highly automated driving. The bin width is 1 s.

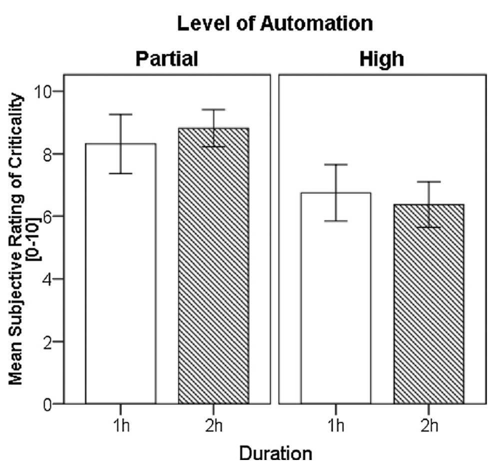

Fig. 15. Comparison of subjective criticality ratings of the target event between the 1h- and 2h-duration for partially and highly automated driving. The error bars represent the 95%-confidence intervals.

Table 8
Effects of subjective observer ratings for fatigue, visual workload, mental workload, motoric workload, and motivation on braking reaction time in partially automated driving.

	Partially automated driving				
			95% - CI		
Predictor	*B*	*SE B*	lower limit	upper limit	*β*
Constant	1.32	0.19	0.92	1.72	
Fatigue	0.53	0.14	0.24	0.82	0.59**
Visual workload	−0.45	0.16	−0.77	−0.12	-0.86**
Mental workload	−0.29	0.13	−0.54	−0.03	−0.44*
Motoric workload	−0.12	0.21	−0.54	0.31	−0.14
Motivation	0.99	0.19	0.59	1.38	1.69***
R^2	.56				
Adjusted R^2	.48				
F	6.46**				

* p < . 05 ** p < .01 ***p < .001.

3.3.1.2. Highly automated driving (System deactivation time). The regression models calculated to predict take-over reaction time in highly automated driving (F(5,26) = 1.79, p = .150, n = 32) and brake reaction times to the braking vehicle were not statistically significant (F(5,26) = 0.85, p = .528, n = 32). In contrast to partially automated driving, there was thus no impact of drowsiness and secondary task engagement on the time it took drivers to take back manual control over the vehicle and react to the braking vehicle.

3.3.2. Predicting minimum TTC

3.3.2.1. Partially automated driving. The regression model calculated to predict minimum TTC (Table 9) in partially automated driving yielded a significant regression equation. As expected, both fatigue (t(26) = −3.34, p = .003) and motivation (t(26) = −5.23, p < .001) contributed significantly to the prediction of the minimal TTC. An increase of one increment in fatigue ratings decreased the TTC_{min}-values by 0.93 s, respectively 1.98 s in motivation ratings. Moreover, mental workload (t(26) = 3.89, p = .001) contributed significantly to predicting minimum TTC. Interestingly, an increase of one increment in mental workload ratings increased minimum TTC by 0.97 s, which can be attributed to the faster brake reaction times found with increased mental workload. Visual (t(26) = 1.53, p = .139) and motoric (t(26) = 1.27, p = .214) workload rating did, however, not significantly contribute to predicting minimal TTC. The partial regression plots of the significant predictors are illustrated in Fig. 17.

3.3.2.2. Highly automated driving. Contrary to expectations, the regression model calculated to predict minimum TTC in highly automated driving did not reach statistical significance (F(5,26) = 0.77, p = .582, n = 32). As with reaction times to the TOR and to the braking vehicle, the criticality of the situations as measured by the TTC was not influenced by fatigue and secondary task engagement in this condition.

3.3.3. Predicting subjective rating of criticality

3.3.3.1. Partially automated driving. The regression model calculated to predict subjective criticality ratings in partially automated driving marginally failed to reach significance (F(5,26) = 2.00, p = .112, n = 32).

3.3.3.2. Highly automated driving. As in partially automated driving, the regression model calculated to predict subjective criticality in highly automated driving did not reach significance (F(5,26) = 0.58, p = .714, n = 32).

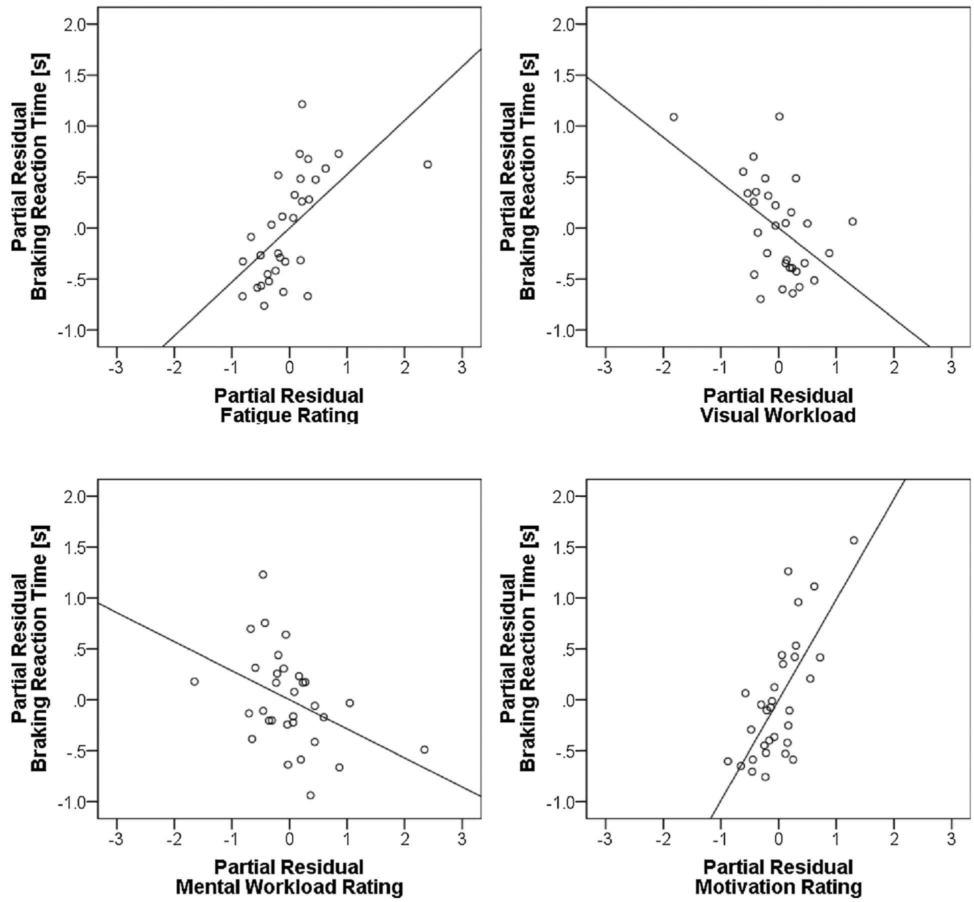

Fig. 16. Partial regression plots for the regression model on braking reaction time in partially automated driving.

Table 9
Effects of subjective observer ratings for fatigue, visual workload, mental workload, motoric workload, and motivation on minimum TTC in partially automated driving.

	Partially automated driving				
			95% - CI		
Predictor	B	SE B	lower limit	upper limit	β
Constant	3.12	0.38	2.33	3.90	
Fatigue	−0.93	0.28	−1.51	−0.36	-0.49**
Visual workload	0.47	0.31	−0.16	1.11	0.44
Mental workload	0.97	0.25	0.45	1.48	0.70**
Motoric workload	0.52	0.41	−0.32	1.36	0.30
Motivation	−1.98	0.38	−2.76	−1.20	-1.63***
R^2	.59				
Adjusted R^2	.51				
F	7.56***				

* $p < .05$ ** $p < .01$ ***$p < .001$.

4. Summary and discussion

The study investigated how engagement in NDRTs and drowsiness level jointly influence drivers' reactions to an imminent lead vehicle braking scenario they had to react to in the transition from partially or highly automated driving to manual vehicle control. Drivers completed an automated drive of either one or two hours before the target event.

At the end of the drives, participants had to react manually to a cut-in scenario, in which a lead vehicle performed a lane change to the participant's lane and initiated a sudden braking maneuver. Especially in the partially automated driving condition this target event was very challenging for the driver and represents a worst-case scenario. It can be assumed that the occurrence of such an unpredictably critical situation coinciding with a silent system failure is a rather unlikely event. Additionally, no collision mitigation systems were implemented which would have initiated an emergency braking in this situation. The high criticality of this situation is also reflected in the drivers' subjective criticality ratings, who rated the situation as "dangerous". This explains the short TTC values and the number of collisions that occurred in this situation.

Video footage of a time interval of 15 s prior to the critical target event was coded using an existing coding scheme for drowsiness assessment and a newly developed scheme for NDRT engagement, which covered aspects that are expected to impair the drivers' availability to take over manual vehicle control (based on Marberger et al., 2017):

- visual workload ratings assessed how strongly a NDRT directs the driver's visual attention away from the road (from permanent monitoring to permanently looking away from the road)
- mental workload ratings assessed the mental load associated with executing the NDRT
- motoric workload ratings assessed the motoric workload needed to interface with the vehicle controls (from having the hands on the steering wheel to turning to the back seat)
- motivation ratings assessed how strongly the raters expected the task would incentivize drivers to continue it

While the fatigue, workload and motivational ratings could be used to predict braking reaction times and TTC values in the partially automated drive, predicting take-over duration and TTC values in the highly automated drive was not possible in a similar fashion. The results of the study have several implications that will be discussed in the following sections.

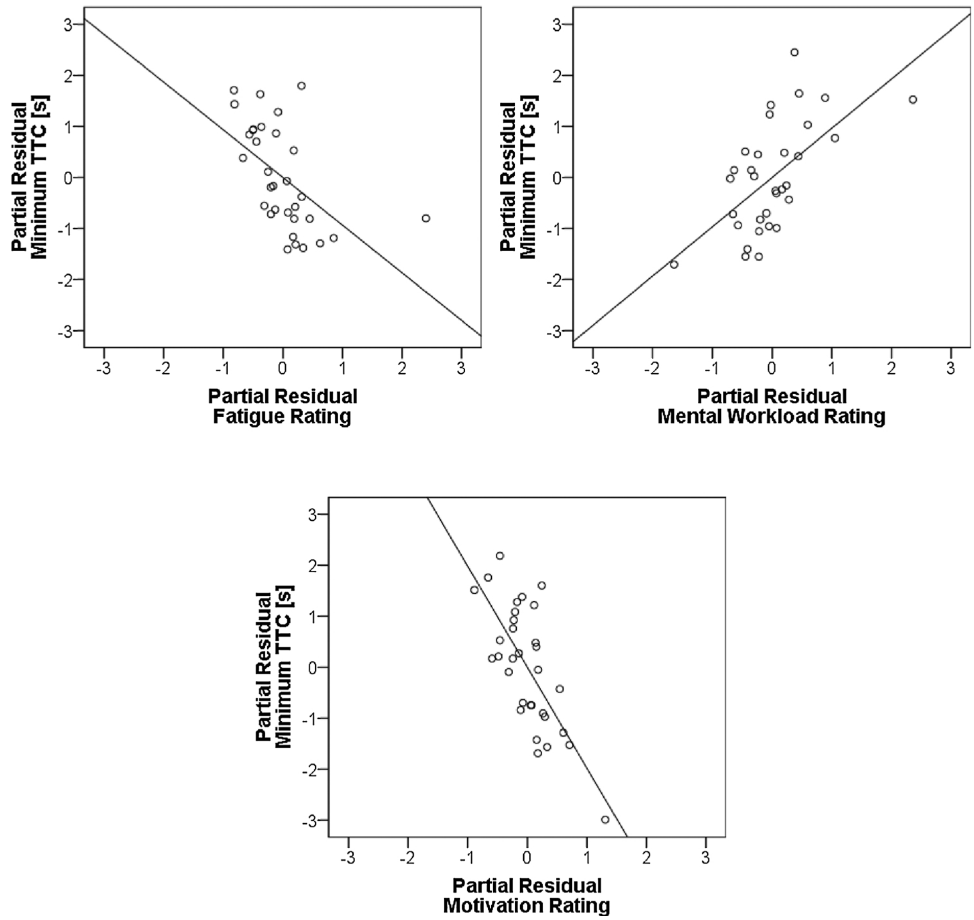

Fig. 17. Partial regression plots for the regression model on minimum TTC in partially automated driving.

4.1. Engagement level in NDRTs and drowsiness development

In both automated driving conditions, the drivers made use of the possibility to carry out activities on their smartphone. Only about 30% of the participants were not engaged in any task at all prior to the target event, regardless of the level of automation. In line with previous research, our findings thus support the view that drivers will readily engage in NDRTs during automated driving (de Winter et al., 2014; Jamson et al., 2013; König and Neumayr, 2017; Large et al., 2017; Naujoks et al., 2016; Naujoks and Totzke, 2014; Pfleging et al., 2016). However, one unsuspected finding from the study is that there were virtually no differences between the automation levels in the observer ratings of workload imposed by these NDRTs. The only exception is the lower motoric workload found in the partially automated conditions, which was due to the instruction that did not allow holding the smartphone in one's hands. It should be noted that the rating scheme developed to assess NDRT engagement was newly developed and firstly used in this study. While the inter-rater agreement demonstrates a good reliability of the method, the lack of difference between the automation levels may also be due to a low sensitivity of the rating scheme. A careful refinement of the definition of the constructs of the rating method (e.g., motivation) and the rating scales used for video coding could therefore be recommended for future studies.

The observed drowsiness level stayed on a relatively low level, which is possibly attributable to the engagement in NDRTs that may have helped drivers to stay within acceptable arousal levels and prevent mental underload (Jarosch et al., 2017; Miller et al., 2015). From these findings it is not surprising that the driver's average take-over times, minimum TTC-values and subjective criticality were not affected by the duration of the study (i.e., there were no statistically significant differences between the one and two hour drives). As opposed to other studies that have found increased drowsiness levels after automated driving periods as short as 15 min (Schömig et al., 2015), drowsiness still stayed on low levels even after prolonged drives of as much as two hours. The results also suggest that the higher event rate experienced by the drivers in the longer drive did not cause significant learning effects, which could have decreased reaction times (Hergeth et al., 2017). While being out of the scope of this study, it might be possible that the participants' behavior during the events prior to the target event could be used to predict their performance. This aspect might be a fruitful research question in future studies.

4.2. Relationship between NDRT engagement, drowsiness and driver performance

4.2.1. Partially automated driving

Regression analysis showed a more detailed picture of the relationship between drowsiness and driver performance in partially automated driving. When required to take back vehicle control during partially automated driving, it was evident that drowsiness led to increased reaction times, as could be expected from prior studies (Neubauer et al., 2014). Engagement with NDRTs further impeded the drivers' reactions when they carried out seemingly (from the observer's point of view) motivating tasks that may have caused them to prioritize executing the NDRT over monitoring the driving environment (Naujoks et al., 2017b; Wickens et al., 2015).

An increase in observed visual and mental workload of performing a NDRT, however, went along with faster reactions to the cutting-in vehicle. This means that participants that were rated to shift their attention between the driving task and the NDRT (opposed to participants that constantly looked on the road) as well as those participants that were rated to engage in more cognitively demanding NDRTs braked faster in reaction to the target event. Cognitive, but not visual workload also increased the TTC-values, which indicates a lower criticality of the situations, but there was no impact on the self-reported situation criticality. These somewhat surprising findings may be due to the fact that the observed visual and mental workload levels usually did not surpass "moderate" workload, which may be just enough to keep the driver on an acceptable arousal level during the automated drive.

For example, Neubauer et al. (2012) also reported faster reactions to a braking event after a control transition from automated to manual driving when drivers were given the possibility to engage in cell phone tasks compared to a control condition in which not tasks were allowed. The authors attributed this finding to increased alertness caused by the engagement in NDRTs compared with the no-task condition that leads to a state of passive fatigue. Gold, Reinder and Bengler (2018) also compared the effects of different NDRTs on take-over performance in highly automated driving and found that an increased cognitive load due to NDRTs decreased the criticality of take-over situations as measured by the TTC. Gold et al. (2018) attributed this effect to an activation of attentional resources or an overcompensation by drivers that are engaged in cognitively demanding NDRTs (e.g., by a stronger brake reaction). However, when looking at the visual workload associated with NDRTs, Gold et al. (2018) did not find any impact on driver's take-over performance. Taken together, the findings of this study implies that performing tasks other than driving, which has clear negative consequences when in manual vehicle control (Purucker et al., 2017a), may not always have negative consequences during partially automated driving.

However, this implication should be treated with great caution as drivers were clearly instructed on their responsibilities (i.e., that they are still fully in charge) and possibilities to use their smartphones were restricted (i.e., drivers had to use the smartphone holder). It was presumably due to the latter that no influence of observed motoric workload on driver performance was found. Drivers could also self-pace the interaction with NDRTs. In a recent study, Pätzold et al. (2017) compared different NDRTs with varying visual workload levels (operationalized by the mean and total eyes-off road time) and found decreased reaction times to system failures such as the ones investigated in this study. In conclusion, the results of the present study clearly point out the need to better understand the consequences of NDRT engagement for the driver's take-over capabilities during partially automated driving. Significantly more research is needed to fully understand how different aspects of NDRTs impair or facilitate the driver's availability to detect and react to failures of partially automated driving systems. For example, the study did not include a reference condition in which no NDRT engagement was allowed at all, which could have produced much larger drowsiness compared to the results observed in this study.

4.2.2. *Highly automated driving*

In comparison to partially automated driving, the regression analyses carried out for the highly automated condition did not reveal any effects. None of the predictors of drowsiness and NDRT engagement were successful in predicting take-over times and subjective criticality of the subsequent target scenario. These findings seemingly contradict earlier studies that have clearly established a link between take-over times and NDRT carried out prior to taking back manual vehicle control (Dogan et al., 2017; Körber et al., 2016; Merat et al., 2012). At least three reasons could account for the current finding. First, due to the instruction in the study, most drivers interacted with a hand-held smartphone. Although different applications were carried out by the drivers, this focus on hand-held smartphones may have inherently limited the variance caused by the NDRT. Other tasks like searching for items in the vehicle interior might lead to a broadened variance and result in more severe impairments of the reaction capabilities. As in partially automated driving, there was also no condition without any NDRT at all, which might have produced a much stronger drowsiness effect.

Second, with a sample size of 32 participants per automated driving condition, the study might not have had enough power to reliably detect an effect associated with the engagement level in the NDRTs. Thus, future studies should replicate the results with a higher sample size, taking the current results as a starting point for a power analysis. Future studies should also use more sophisticated statistical procedures that can lead to better modelling results. For example, in prediction of the subjective criticality, using an ordered logistic regression model would have been more suitable to predict ordinal data (e.g., Purucker et al., 2014, 2017b). Limiting the model predictions can also lead to more precise results (e.g., a reaction time cannot be lower than zero).

Third, the HMI concept used in the study allowed a self-paced disengagement from the NDRT and take-over of manual control, but also involved a timer. It may be possible that this HMI concept caused a convergence of take-over times as drivers may not have felt the pressure to immediately react to the TOR, but also may have avoided reacting lately to it. While the braking reaction times in the partially automated condition are likely to be manual take-overs that were performed "as-fast-as-possible reactions" (therefore being severely impacted by the driver's current take-over capability state), take-over times in highly automated driving in not so time-critical situations might be more likely influenced by HMI design or the driver's personal preference. This assumption is in line with previous research indicating that the impact of NDRT engagement on take-over time is dependent on the driver's perceived urgency to intervene (Zeeb et al., 2016). As for the criticality of the driving situation, no carry-over effect from the preceding 15s-interval was found. It is possible that the drivers successfully managed to come back into the control loop before deactivating the automation and reclaiming control over the vehicle. During the transition period, the automation still carried out the driving task completely, and the critical driving event only happened after the drivers themselves had requested and successfully taken over control. On a more general note, the study results on manual take-over from highly automated driving are likely not transferable to situations with a higher time pressure that do not allow the "driver transition process" (Marberger et al., 2017) to be completed before the transfer of control happens.

5. Conclusions

- Drivers readily engaged in various NDRTs on their smartphones during a one or two hours partially or highly automated drive.
- In spite of a rather long automated drive, drowsiness stayed on a relatively low level, presumably as a result of stimulating interactions with the smartphone.
- The duration of the automated drive did not influence the driver's take-over performance, neither during partially nor during highly automated driving.
- In partially automated driving, the individual drowsiness levels and observed motivational appeal of NDRTs had a negative impact on driver's performance, while low to moderate levels of visual and mental workload increased driver's performance in a worst-case driving scenario.
- In highly automated driving, no carry-over effects from NDRTs or drowsiness level on driver's braking performance were observed, which may be explained by the experimental setup that provided drivers with a sufficiently long time budget to prepare for taking over control and allowed drivers to re-engage in the driving task at their own pace.

Acknowledgements

This study was sponsored by the Robert Bosch GmbH. The Human-Machine Interface used in this study was an experimental prototype that is not intended for series production. The views expressed in this article are the ones of the authors.

Appendix A. Assumptions of multiple regression

Before using multiple regressions, there is a need to check data for prerequisites. One prerequisite is that the residuals of the criterion variables are normally distributed. Data was screened for normal distribution using the Kolmogorov-Smirnov-test.

As depicted in Table A1, the Kolmogorow-Smirnov test indicated that subjective criticality ratings in partially automated driving, system deactivation time in highly automated driving and minimum TTC in highly automated driving were not normally distributed. Analyzing the histograms of standardized residuals, however, showed that there was no extreme deviation from normality. As it is known that the Kolmogorow-Smirnow test is very sensitive towards outliers (Keller, 2014) and that "only extreme departures of the distribution […] from normality yield spurious results [in regression analysis]" (Kleinbaum, Kupper, Muller & Nizam, 1998, p .117), it was refrained from transforming data.

Another prerequisite is that there is no multicollinearity. Therefore, the correlations between predictors (Table A2) should be considered as they provide initial indication of multicollinearity. According to Field (2005), there should be no correlations that are higher than $r = .90$. All correlations were below this threshold. Examining the variance inflation factor (VIF) did also not indicate multicollinearity, as all VIF-values are below 10 (Myers 1990, see Table A3).

Table A1
Results of the Kolmogorow-Smirnov Test.

	Variable	*N*	Kolmogorow-Smirnov (*p*)
Partially automated driving	Braking reaction time	31	.20
	Minimum TTC	32	.20
	Subjective criticality	32	.00
Highly automated driving	System deactivation time	32	.01
	Minimum TTC	32	.00
	Subjective criticality	32	.09

Table A2
Correlations between rating dimensions.

	Fatigue	Visual workload	Mental workload	Motoric workload	Motivation
Fatigue	1	−.42*	−.37*	−.45**	−.27
Visual workload		1	.63***	.77***	.87***
Mental workload			1	.64***	.70***
Motoric workload				1	.63***
Motivation					1

* $p < .05$ ** $p < .01$ *** $p < .001$.

Table A3
Overview of variance inflation factors of predictors in the different regression models.

	Regression models: Partially automated driving			Regression models: Highly automated driving		
VIF Predictor	*Braking reaction time*	*Minimum TTC*	*Subjective criticality*	*System deactivation time*	*Minimum TTC*	*Subjective criticality*
Fatigue	1.39	1.39	1.39	1.35	1.35	1.35
Visual workload	5.17	5.29	5.29	6.61	6.61	6.61
Mental workload	2.07	2.08	2.08	2.35	2.35	2.35
Motoric workload	3.50	3.51	3.51	3.00	3.00	3.00
Motivation	6.11	6.23	6.23	5.30	5.34	5.34

References

Altmann, E.M., Trafton, J.G., 2004. Task interruption: resumption lag and the role of cues. Proc. Ann. Meeting Cogn. Sci. Soc. 26, 43–48.

Braunagel, C., Geisler, D., Rosenstiel, W., Kasneci, E., 2017a. Online recognition of driver-activity based on visual scanpath classification. IEEE Intell. Transp. Syst. Mag. 9 (4), 23–36.

Braunagel, C., Rosenstiel, W., Kasneci, E., 2017b. Ready for take-over? A new driver assistance system for an automated classification of driver take-over readiness. IEEE Intell. Transp. Syst. Mag. 9 (4), 10–22.

de Winter, J., Happee, R., Martens, M.H., Stanton, N.A., 2014. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: a review of the empirical evidence. Transp. Res. Part F 27B, 196–217.

Desmond, P., Hancock, P., Monette, J., 1998. Fatigue and automation-induced impairments in simulated driving performance. Trans. Res. Rec. J. Transp. Res. Board (1628), 8–14.

Dogan, E., Rahal, M.-C., Deborne, R., Delhomme, P., Kemeny, A., Perrin, J., 2017. Transition of control in a partially automated vehicle: effects of anticipation and non-driving-related task involvement. Transp. Res. Part F 46, 205–215.

Eriksson, A., Stanton, N.A., 2017. Take-over time in highly automated vehicles: non-critical transitions to and from manual control. Hum. Factors 59 (4), 689–705.

Fleiss, J.L., Cohen, J., 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ. Psychol. Meas. 33 (3), 613–619.
Gasser, T.M., Arzt, C., Ayoubi, M., Bartels, A., Bürkle, L., Eier, J., et al., 2012. Legal consequences of an increase in vehicle automation. Die Bundesanstalt für Straßenwesen (BASt) Report F 83.
Gold, C., Damböck, D., Lorenz, L., Bengler, K., 2013. "Take over!" How long does it take to get the driver back into the loop? Proceedings of the Human Factors and Ergonomics Society Annual Meeting 57, 1938–1942.
Gold, C., Berisha, I., Bengler, K., 2015. Utilization of drivetime–performing non-driving related tasks while driving highly automated. Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
Gold, C., Happee, R., Bengler, K., 2018. Modeling take-over performance in level 3 conditionally automated vehicles. Accid. Anal. Prev. 116, 3–13.
Hayward, J.C., 1972. Near miss determination through use of a scale of danger. Highway Res. Rec. 384, 24–34.
Hergeth, S., Lorenz, L., Krems, J., 2017. Prior familiarization with takeover requests affects drivers' takeover performance and automation trust. Hum. Factors 59 (3), 457–470.
Jamson, A.H., Merat, N., Carsten, O.M., Lai, F.C., 2013. Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. Transp. Res. Part C 30, 116–125.
Jarosch, O., Kuhnt, M., Paradies, S., Bengler, K., 2017. It's out of our hands now! Effects of non-driving related tasks during highly automated driving on drivers' fatigue. 9th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.
Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A.M., et al., 2010. Control and interference in task switching—a review. Psychol. Bull. 136 (5), 849.
Ko, S.M., Ji, Y.G., 2018. How we can measure the non-driving-task engagement in automated driving: comparing flow experience and workload. Appl. Ergon. 67, 237–245.
König, M., Neumayr, L., 2017. Users' resistance towards radical innovations: the case of the self-driving car. Transp. Res. Part F 44, 42–52.
Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med. 15 (2), 155–163.
Körber, M., Gold, C., Lechner, D., Bengler, K., 2016. The influence of age on the take-over of vehicle control in highly automated driving. Res. Part F Traffic Psychol. Behav. 39, 19–32.
Large, D.R., Burnett, G., Morris, A., Muthumani, A., Matthias, R., 2017. A longitudinal simulator study to explore drivers' behaviour during highly-automated driving. Paper Presented at the International Conference on Applied Human Factors and Ergonomics (AHFE 2017).
Louw, T., Madigan, R., Carsten, O., Merat, N., 2016. Were they in the loop during automated driving? Links between visual attention and crash potential. Inj. Prev injuryprev-2016-042155.
Marberger, C., Mielenz, H., Naujoks, F., Radlmayr, J., Bengler, K., Wandtner, B., 2017. Los Angeles, USADriver Availability Paper Presented at the 8th International Conference on Applied Human Factors and Ergonomics (AHFE 2017)2017. Driver Availability Paper Presented at the 8th International Conference on Applied Human Factors and Ergonomics (AHFE 2017).
Merat, N., Jamson, A.H., Lai, F.C., Carsten, O., 2012. Highly automated driving, secondary task performance, and driver state. Hum. Factors 54 (5), 762–771.
Miller, D., Sun, A., Johns, M., Ive, H., Sirkin, D., Aich, S., et al., 2015. Distraction becomes engagement in automated driving. Proc. Hum. Factors Ergon. Soc. Ann. Meeting 59 (1), 1676–1680.
Naujoks, F., Totzke, I., 2014. Behavioral adaptation caused by predictive warning systems–the case of congestion tail warnings. Transp. Res. Part F 26, 49–61.
Naujoks, F., Mai, C., Neukum, A., 2014. The effect of urgency of take-Over requests during highly automated driving under distraction conditions. In: In: Ahram, T., Jang, R.-L. (Eds.), Proceedings of the 5th AHFE Conference Vol. 7. AHFE Conference, Kraków Poland, pp. 2099–2106.
Naujoks, F., Purucker, C., Neukum, A., Wolter, S., Steiger, R., 2015. Controllability of partially automated driving functions–does it matter whether drivers are allowed to take their hands off the steering wheel? Transp. Res. Part F 35, 185–198.
Naujoks, F., Purucker, C., Neukum, A., 2016. Secondary task engagement and vehicle automation–comparing the effects of different automation levels in an on-road experiment. Transp. Res. Part F 38, 67–82.
Naujoks, F., Purucker, C., Wiedemann, K., Neukum, A., Wolter, S., Steiger, R., 2017a. Driving performance at lateral system limits during partially automated driving. Accid. Anal. Prev. 108, 147–162.
Naujoks, F., Wiedemann, K., Schömig, N., 2017b. The importance of interruption management for usefulness and acceptance of automated driving. Paper Presented at the 9th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg.
Naujoks, F., Befelein, D., Wiedemann, K., Neukum, A., 2018. A review of Non-driving-related tasks used in studies on automated driving. In: Stanton, N. (Ed.), Advances in Human Aspects of Transportation (AHFE 2017). Springer, Cham, pp. 525–537.
Neubauer, C., Matthews, G., Saxby, D., 2012. The effects of cell phone use and automation on driver performance and subjective state in simulated driving. Paper Presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
Neubauer, C., Matthews, G., Saxby, D., 2014. Fatigue in the automated vehicle: do games and conversation distract or energize the driver? Proc. Hum. Factors Ergon. Soc. Ann. Meeting 58 (1), 2053–2057.
Neukum, A., Krüger, H.-P., 2003. Fahrerreaktionen bei Lenksystemstörungen–Untersuchungsmethodik und Bewertungskriterien. VDI-Berichte 1791, 297–318.
Pätzold, A., Schmidt, C., Rauh, N., Cocron, P., Hergeth, S., Keinath, A., et al., 2017. From distraction to controlled engagement: secondary tasks affect drivers' supervisory and fall-back performance while using SAE level 2 driving automation. HFES Europe Chapter Annual Meeting.
Pfleging, B., Rang, M., Broy, N., 2016. Investigating user needs for non-driving-related activities during automated driving. 12.-15.12.2016. Paper Presented at the Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia (MUM' 16), Rovaniemi, Finland.
Purucker, C., Rüger, F., Schneider, N., Neukum, A., Färber, B., 2014. Comparing the perception of critical longitudinal distances between dynamic driving simulation, test track and vehicle in the loop. In: Stanton, N., Landry, S., Di Bucchianico, G., Vallicelli, A. (Eds.), Advances in Human Aspects of Transportation: Part I, pp. 421–430 Danvers: AHFE Conference.
Purucker, C., Naujoks, F., Prill, A., Neukum, A., 2017a. Evaluating distraction of in-vehicle information systems while driving by predicting total eyes-off-road times with keystroke level modeling. Appl. Ergon. 58, 543–554.
Purucker, C., Schneider, N., Rüger, F., Frey, A., 2017b. Validity of research environments. Comparing criticality perceptions across research environments. In: Bengler, K., Hoffmann, S., Manstetten, D., Neukum, A., Drüke (Hrsg.), J. (Eds.), UR:BAN Human Factors in Traffic Approaches for Safe Efficient and Stressfree Urban Traffic. Springer Vieweg, Wiesbaden, Germany.
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. Proc. Hum. Factors Ergon. Soc. Ann. Meeting 58 (1), 2063–2067.
Schömig, N., Hargutt, V., Neukum, A., Petermann-Stock, I., Othersen, I., 2015. The interaction between highly automated driving and the development of drowsiness. Procedia Manuf. 3, 6652–6659.
Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86 (2), 420.
Trafton, J.G., Altmann, E.M., Brock, D.P., Mintz, F.E., 2003. Preparing to resume an interrupted task: effects of prospective goal encoding and retrospective rehearsal. Int. J. Hum. Stud. 58 (5), 583–603.
Trivedi, M.M., Gandhi, T., McCall, J., 2007. Looking-in and looking-out of a vehicle: computer-vision-based enhanced vehicle safety. IEEE Trans. Intell. Transp. Syst. 8 (1), 108–120.
Vogelpohl, T., Vollrath, M., 2017. Müdigkeit Und Hochautomatisiertes Fahren. GDV, Berlin.
Wickens, C.D., Gutzwiller, R.S., Santamaria, A., 2015. Discrete task switching in overload: a meta-analyses and a model. Int. J. Hum. Stud. 79, 79–84.
Wiegand, D.M., McClafferty, J., McDonald, S.E., Hanowski, R.J., 2009. Development and Evaluation of a Naturalistic Observer Rating of Drowsiness Protocol. Virginia Tech Transportation Institute, Virginia Tech.
Wierwille, W.W., Ellsworth, L.A., 1994. Evaluation of driver drowsiness by trained raters. Accid. Anal. Prev. 26 (5), 571–581.
Zeeb, K., Buchner, A., Schrauf, M., 2016. Is take-over time all that matters? The impact of visual-cognitive load on driver take-over quality after conditionally automated driving. Accid. Anal. Prev. 92, 230–239.
Zeeb, K., Härtel, M., Buchner, A., Schrauf, M., 2017. Why is steering not the same as braking? The impact of non-driving related tasks on lateral and longitudinal driver interventions during conditionally automated driving. Transp. Res. Part F Traffic Psychol. Behav. 50, 65–79.