Taking over vehicle control in time-critical and dynamic situations after automated driving

vorgelegt von

M. Sc.

Fabienne Rahel Roche

ORCID: 0000-0003-4875-8622

an der Fakultät V – Verkehrs- und Maschinensysteme

der Technischen Universität Berlin

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

– Dr. rer. nat. –

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Dietrich Manzey

Gutachter: Prof. Dr. Manfred Thüring

Gutachter: Prof. Dr. Martin Baumann

Tag der wissenschaftlichen Aussprache: 14. Juli 2022

Berlin, 2022

Acknowledgments

Over the past years of researching and completing the manuscript, I have received great support from different persons. I would like to take the opportunity to thank them, first of all, my supervisor Manfred Thüring. You were always open to discussing experimental settings, testing our scenarios in the driving simulator for the umpteenth time, and accompanying me through the process. Thank you, Martin Baumann, to take the duty to be my second supervisor, to discuss options for the statistical analyses, and to point at aspects in my thesis that needed consolidation.

Not only the professional input was important for the creation of this work but also the working atmosphere at the chair of Cognitive Psychology and Cognitive Ergonomics, Technische Universität Berlin. Christin Böschow, Stefan Brandenburg, Sandra Becker, Katharina Jungnickel, Friederice Kuhn, Michael Minge, Nikolaus Rötting, and Anna Trukenbrod made the office to a place I looked forward going to. Special thanks go to Christin, who knows the bureaucracy of the Technische Universität Berlin better than anyone else and sometimes made impossible possible, to Stefan, who introduced me to the academic world and spent hours reviewing my manuscripts and discussing with me, to Anna, who is an absolute LMM expert and enriched the present thesis with her knowledge, and to Sandra, who was a great partner for solving the problems we faced during the research project. In addition, Steffen Müller and Thang Nguyen from the chair of Automotive Engineering, Technische Universität Berlin, enriched our work with their knowledge and set up the driving simulator, an essential prerequisite for the experimental studies. The Deutsche Forschungsgemeinschaft (DFG, grant number 326727090) funded the research project in which the studies were carried out.

Finally, I would like to thank my family. Waldemar Roche, your faith in me makes me a stronger person, and to believe in me and my skills. You encouraged me to do what I wanted. My greatest thanks go to my parents, who raised me to ask questions, build the basis to be confident that I can reach anything I want, and supported my wish to graduate. Particularly during the last year, they did not only provide motivational but practical support by taking care of my son many times while I worked on the publications and the final manuscript. Your energy and your dedication to your children (and grandchild) are remarkable.

Abstract

In the past years, automated driving has been receiving much attention from researchers and engineers around the world. As a result, various advanced driver assistance systems are installed as standard or are far developed. However, they may still fail and request the driver to take back vehicle control. Moreover, drivers are allowed to override the system whenever they wish. These driver-initiated take-overs may happen anytime, even in driving situations with low time budgets and high vehicle dynamics. The conditions present challenges since fast and appropriate take-overs are required while the drivers have little time to react, face dynamic driving situations, and potentially hardly have experience and routine in taking over. To date, driver-initiated take-overs have barely been investigated. Therefore, it is questionable how drivers react and how they perform. Hence, the present thesis investigates *what might happen if drivers take over control in time-critical and dynamic driving situations.* This question includes behavior and performance but also the perceived criticality.

In the first driving simulator study, two criticality assessment tools were tested on validity. Participants (N_1 = 25) experienced take-over situations with different time budgets and evaluated the criticality on two rating scales. The results showed that they are both equally valid and that increasing practice neither changed the perceived effort ratings nor the take-over behavior. Except braking became weaker with more experience. In the second and third driving simulator study (N_2 = 42, N_3 = 60), it was investigated how drivers perceive time-critical and dynamic take-over situations and how they behave and perform when taking over. The participants were triggered to take back control in brake (study 2) and double lane change situations (study 3). The results demonstrated that drivers could differentiate between different degrees of objective criticality and adapted their braking and steering to the driving situation. Take-over times were very low under all conditions and were hardly affected by the driving situations. Drivers decelerated and steered stronger than necessary and changed lanes without reason. Thereby, they risked vehicle instability, lane departures, rear-end collisions, and collisions with overtaking vehicles. This indicates that taking over in time-critical and dynamic driving situations is hazardous. To avoid or mitigate the consequences of inappropriate reactions, drivers could be supported by an assistance system that modifies their input. Two of these versions are discussed.

Keywords: Automated Driving, Behavioral Change, Criticality Assessment, Critical Driving Situations, Driver Behavior, Driver-Vehicle Interaction, Driving Simulator, Scale Validation, Take-Over, Vehicle Automation

Table of contents

List of figures

List of tables

List of abbreviations

ADS	Automated driving system
CRS	Criticality Rating Scale
DDT	Dynamic driving task
M	Mean
NDRT	Non-driving related task
ODD	Operational design domain
OEDR	Object and event detection and response
RQ	Research question
SAE	Society of automotive engineers
SD	Standard deviation
SCA	Scale of Criticality Assessment of Driving Situations
THW	Time headway
TOR	Take-over request
TOT	Take-over time
TTC	Time-to-collision
TU	Traction usage

1 Introduction

Automated driving promises to enrich 'lives around the world' (Toyota Automated Driving, 2020, p. 3). This statement is based on several advantages that automated driving implies. The first advantage is that everyday traffic is supposed to become safer with increasing share of automated vehicles on the road. Worldwide, over 1.3 million people die every year in road traffic accidents of which 75 % can be attributed to driver error, e.g. distraction or speeding (Destatis, 2021; European Commission, 2016; World Health Organization, 2021). These numbers show that there is a high need to reduce the impact of the human factor to increase safety. Automated driving may achieve this by taking over parts of the drivers' task (Anderson et al., 2014; SAE International, 2018).

Second, road transport is expected to become less time- and energy-consuming with automated driving systems (ADS). From mid 2019 to mid 2020, the average U.S. American driver spent up to 59 minutes per day travelling in the car, with rising tendency over the past years (AAA Foundation for Traffic Safety, 2021). Driving automated, they can invest this time in other activities and, thereby, be more productive. Besides, car-to-x-communication and platoon traveling will increase the effective velocities, decrease the peak speeds, and reduce congestions which result in shorter travel times and less energy consumption (Anderson et al., 2014; Department for Transport, 2015; Kesting et al., 2008).

Third, vehicles with ADS are expected to be more inclusive than manually operated ones. Individuals who are unable or have difficulties to drive manually operated vehicles due to their physical state, e.g. disabled or elderly people, are currently mostly excluded from the benefits of this type of individual mobility. Automated driving opens up the access for most of them resulting in a more self-determined, independent life (Anderson et al., 2014; Casner et al., 2016; Department for Transport, 2015; ERTRAC, 2015; Fagnant & Kockelman, 2015).

For these reasons, it is not striking that automated driving is one of the most up-to-date topics of the automotive industry (BMW, 2020; General Motors, 2020; Hyundai, 2020; Toyota Automated Driving, 2020; VW, 2020).

1.1 Definitions of driving automation

Vehicles can be automated to different extents depending on the number and type of subtasks they control. Therefore, the implementations of driving automation can be divided into levels. Different definitions were introduced by the German Federal Highway Research Institute (BASt) (Gasser et al., 2012), the US-American National Highway Traffic Safety

Administration (U.S. Department of Transportation, National Highway Traffic Safety Administration, 2013), and the International Society of Automotive Engineers (SAE International, 2021). One problem with the definitions is that they use different terms to describe the same levels and similar words to relate to distinct levels, which may lead to misunderstandings. In the present thesis, the SAE-definition will be used because it is standard in human factors research and the automotive industry (Eriksson & Stanton, 2017; General Motors, 2020; Hyundai, 2020; Toyota Automated Driving, 2020).

The SAE-classification is based on two aspects: the allocation of the dynamic driving task (DDT) and the conditions in which the automated system can operate, the operational design domain (ODD, see Figure 1; SAE International, 2021). The DDT consists of the lateral and longitudinal vehicle control and the object and event detection and response (OEDR). The latter describes monitoring and scanning the environment. Based on this task allocation, six different levels of driving automation are identified. Figure 1 presents the distinction including a short description. In the present thesis, driving automation systems at SAE-level 3 are investigated, hence only this one is presented in more detail:

At SAE-level 3, *Conditional Driving Automation*, the ADS controls lateral and longitudinal vehicle motion and performs the OEDR. This capability is limited to specific driving modes (ODD), and the drivers are expected to be the fallback when the ADS reaches its system limits. In this case, a take-over request (TOR) informs the driver about the urge to take back vehicle control.

Even though fully automated driving will solve diverse human factor issues, the transition is challenging. Lower levels of ADS will be introduced in the market before higher ones are technically feasible and legally permitted (Noy et al., 2018). The ADS with reduced capabilities may be hazardous because it might still request the drivers to take back control when reaching its system limits (Casner et al., 2016; SAE International, 2021). Moreover, based on the amendments of the Vienna Convention on Road Traffic from 1968, drivers are allowed to regain vehicle control whenever they wish to do so (United Nations Economic Commission for Europe, 2014). The present thesis focuses on the latter type of take-overs. Both types shall be presented briefly in the following, supplemented by the criticality of the driving situation. They can be summarized to the characteristics of take-overs.

Level	Name	Narrative definition	DDT		DDT Fallback	ODD
			Sustained lateral and longitudinal vehicle motion control	OEDR		
Driver performs part or all of the DDT						
0	No Driving Automation	The performance by the driver of the entire DDT, even when enhanced by active safety systems.	Driver	Driver	Driver	n/a
1	Driver Assistance	The sustained and ODD-specific execution by a driving automation system of either the lateral or the longitudinal vehicle motion control subtask of the DDT (but not both simultaneously) with the expectation that the driver performs the remainder of the DDT.	Driver and System	Driver	Driver	Limited
2	Partial Driving Automation	The sustained and ODD-specific execution by a driving automation system of both the lateral and longitudinal vehicle motion control subtasks of the DDT with the expectation that the driver completes the OEDR subtask and supervises the driving automation system.	System	Driver	Driver	Limited
ADS performs the entire DDT (while engaged)						
3	Conditional Driving Automation	The sustained and ODD-specific performance by an ADS of the entire DDT with the expectation that the DDT fallback-ready user is receptive to ADS-issued requests to intervene, as well as to DDT performance relevant system failures in other vehicle systems, and will respond appropriately.	System	System	Fallback-ready user	Limited
4	High Driving Automation	The sustained and ODD-specific performance by an ADS of the entire DDT and DDT fallback without any expectation that a user will respond to a request to intervene.	System	System	System	Limited
5	Full Driving Automation	The sustained and unconditional (i.e., not ODD-specific) performance by an ADS of the entire DDT and DDT fallback without any expectation that a user will respond to a request to intervene.	System	System	System	Unlimited

Figure 1: SAE-levels of driving automation, own representation based on SAE International (2018).

Note. ADS: automated driving system; DDT: dynamic driving task; OEDR: object and event detection and response; ODD: operational design domain.

1.2 Characteristics of take-overs

Take-overs can be classified based on the initiator of the transition (the system or the driver) and the criticality of the driving situation, as proposed by McCall et al. (2016). The authors used a binary classification for criticality. To allow for a finer gradation and to better reflect the reality, it is here considered continuous ranging from less to more critical. McCall et al. (2016) also suggested differentiating whether the take-over is scheduled or not. This aspect is not relevant for the present thesis; hence, I do not refer to it. The distinction is visualized in Figure 2 and presented in more detail in the following.

1.2.1 Initiator of take-overs

Take-overs can be *system-initiated* (also referred to as *automation-initiated* or *passive disengagements*) or *driver-initiated* (also called *operator-initiated*, *user-initiated,* or *active disengagement*; see Figure 2). In the former ones, the ADS initiates the transition of control by a TOR.

In driver-initiated take-overs, the drivers decide to override the ADS (Lu et al., 2016; Maggi et al., 2020; Martens et al., 2007; McCall et al., 2016; Melcher et al., 2015). The amendments of the Vienna Convention paved the way for this kind of take-overs. It stated that ADS are allowed if they ‘can be overridden or switched off by the driver‘ in any driving situation (United Nations Economic Commission for Europe, 2014, p. 9). There might be various reasons for driver-initiated take-overs: drivers are frightened by a certain event, they doubt that the ADS can handle the situation, they do not understand what the ADS is doing, or they want to experience the joy of driving (Lu et al., 2016; McCall et al., 2016).

1.2.2 Criticality of driving situations

Driving situations in which take-overs occur, can also be characterized by their criticality (see Figure 2). It expresses the threat or danger of the situation (Herrmann et al., 2015). Based on the ISO 31000 (2009), Rodemerk et al. (2012) understood criticality as the combination of the likelihood of potentially harmful events and the severity of any damage. Hence, an inevitable accident with severe consequences constitutes the highest criticality. In the present thesis, I refer to this holistic definition.

Different parameters may determine the criticality of driving situations. These are, for example, time budget and vehicle dynamics (Junietz et al., 2017).

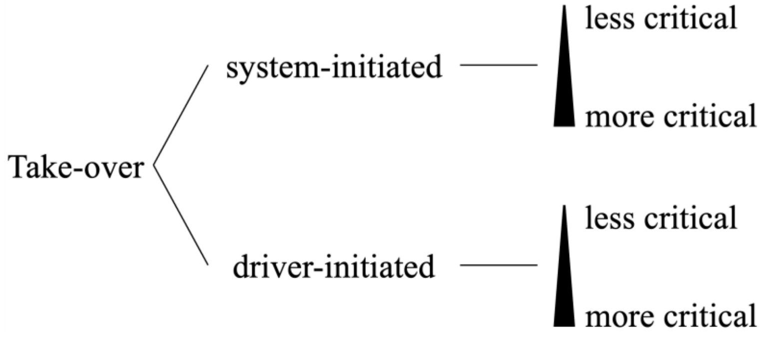

Figure 2: Characteristics of take-overs.
Note. The figure is based on the classification proposed by McCall et al. (2016).

Time budget

Time budget describes the available time for a take-over and is frequently used to vary and characterize driving situations (Damböck et al., 2012; Mok, Johns, Lee, Ive, et al., 2015; Roche & Brandenburg, 2018, 2020; Rodemerk et al., 2012; van den Beukel & van der Voort, 2013). It can be quantified, for example, by the variables time headway (THW) and time-to-collision (TTC), e.g. at the moment of the take-over. THW is defined as the time it would take the ego vehicle to reach the current position of a reference object assuming that they are on the same trajectory and that the ego vehicles' velocity remains constant (Vogel, 2003). A reference object may be a vehicle travelling ahead. THW is computed by the quotient of the distance to the reference object $d_{reference}$ and the velocity of the ego vehicle v_{ego} (see Equation 1). TTC describes the time starting from a certain moment, e,g. a TOR, until the ego vehicle would collide with another object. Equation 2 shows that TTC also takes the velocity of the other object $v_{reference}$ into account.

$$THW = \frac{d_{reference}}{v_{ego}}$$ Equation 1

$$TTC = \frac{d_{reference}}{v_{ego} - v_{reference}}$$ Equation 2

Both parameters, THW and TTC, are respectively more appropriate to describe the criticality of different driving situations. THW is frequently used for car-following scenarios (Eick & Debus, 2005; Siebert et al., 2014) and serves as an indicator for tailgating (Vogel, 2003). The reason is that in these situations, THW is relatively low and depicts the criticality suitable. At the same time, TTC may be very large or even undefinable when $v_{reference}$ is equal to or higher than v_{ego} (Vogel, 2003). In contrast, TTC is often used to characterize the criticality of driving scenarios in which a collision with another vehicle or obstacle is imminent (Gold et al., 2013). In such cases, the velocity of the reference object is zero, hence, THW and TTC are the same. Referring to THW in both scenarios would not consider that the velocities of the reference objects vary strongly. Similar absolute THW-values would describe very different criticalities. Whereas, referring to TTC in both scenarios is also problematic because it is likely to be undefined in car-following ones. This shows that, depending on the situation, one time budget-variable is more appropriate than the other and that a distinction is crucial. Hence, in this thesis, THW is used to describe car-following scenarios, while TTC is referred to in situations with a stationary obstacle.

In driving situations with low THW or TTC, the likelihood of collisions is high because drivers have less time to avoid an impending collision (ISO 31000:2009, 2009). Besides, take-over behavior may be poor because less time is available to gain situation awareness and potentially fewer options left to react. Thus, a worse take-over behavior with severe consequences is more likely (Scott & Gray, 2008). Hence, time budget comprises the likelihood and severity of an accident and is a suitable parameter to describe the criticality of driving situations, as defined by Rodemerk (2012). Based on this, low time budgets can be considered as more critical, while situations with higher time budgets are less critical. A threshold to distinguish between less and more critical is not available. The SAE (2021) only states that ADSs should be designed to request the drivers with a '[…] sufficient time for the fallback-ready user" (p. 31) to take over.

Vehicle dynamics

In contrast to time budget, the situational parameter *vehicle dynamics* is a rather novel variable in human factors research. Different variables can be taken at hand to describe them, depending on the direction and purpose. This thesis focuses on traction usage. Its meaning and relation to vehicle dynamics are explained in the following.

Vehicle dynamics are affected by forces and motions in different directions: longitudinal, lateral, and vertical (Breuer & Rohrbach-Kerl, 2015). The forces in lateral and longitudinal directions (F_{lat} and F_{long}) are influenced by acceleration, braking, and steering. They can be summarized to a resulting horizontal force $F_{horizont}$ (Rajamani, 2011):

$$F_{horizont} = \sqrt{F_{lat}^2 + F_{long}^2} \qquad \text{Equation 3}$$

The *Kamm's circle* (also called the *circle of forces* or *friction ellipse*) presented in Figure 3 visualizes an idealized simplification of the interaction of lateral, longitudinal, resulting horizontal force, and the maximal force that can be transferred from tire to road (Breuer & Rohrbach-Kerl, 2015; Kamm, 1936). Figure 3 shows that the lateral force that can be transferred is lower when longitudinal forces act at the same time. The maximal horizontal force is determined by the vertical forces F_{vert} and the traction coefficient μ, see Equation 4 (Breuer & Rohrbach-Kerl, 2015; Pacejka, 2006; Rajamani, 2011).

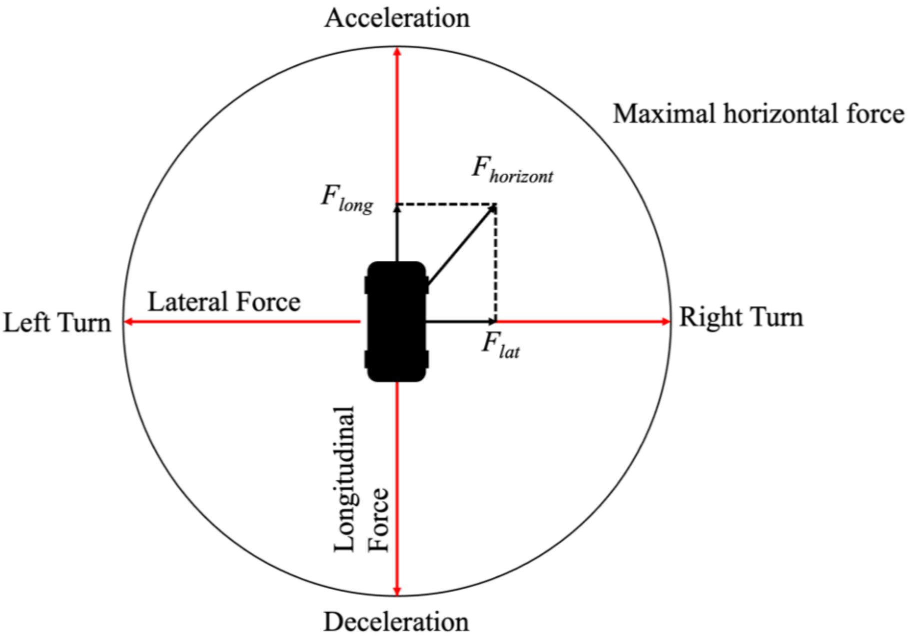

Figure 3: Kamm’s circle based on Kamm (1936).

Note. The depicted vehicle in the middle is heading upwards. The directions of the longitudinal and lateral forces add up to a right curve.

$$\max(F_{horizont}) \leq |\mu * F_{vert}| \qquad \text{Equation 4}$$

As long as the horizontal force $F_{horizont}$ does not exceed the maximal horizontal force, the vehicle is stable (Kamm, 1936; Pacejka, 2006). Exceeding the maximal horizontal force results in slip and loss of vehicle control (Breuer & Rohrbach-Kerl, 2015). The effective horizontal force $F_{horizont}$ is always lower or equals the maximal horizontal force. Based on Equation 3 and Equation 4, this results in the following constraint:

$$\sqrt{F_{lat}^2 + F_{long}^2} \leq \max(F_{horizont}) \leq |\mu * F_{vert}| \qquad \text{Equation 5}$$

The ratio of the effective and the maximal horizontal force describes the stability of the vehicle on the road and is termed traction usage (TU; see Equation 6).

$$TU = \frac{F_{horizont}}{\max(F_{horizont})} = \frac{\sqrt{F_{lat}^2 + F_{long}^2}}{\mu * F_{vert}} \qquad \text{Equation 6}$$

The maximal possible value of TU is 1. A high TU describes a driving situation with strong longitudinal and/or lateral acting forces, e.g. de- or accelerations (Rajamani, 2011). In those situations, a loss of vehicle control is more likely to occur in case an additional acceleration in either direction occurs. Then, the limit of Kamm's circle may be exceeded. This can happen both when driving straight ahead or curves. Thus, the stability of the vehicle gets endangered. The consequences of collisions are more severe because the impact of a collision is worse. Hence, TU can be used to describe the criticality of driving situations (Rodemerk et al., 2012). High TU characterizes more critical situations, while low TU can be considered as less critical.

1.2.3 Conclusion

Take-overs can be system- or driver-initiated. In both cases, the driving situations may be critical depending on – among other – time budget and vehicle dynamics. These situational parameters might influence the subjective experience, behavior, and performance after initiating a take-over.

1.3 Research on take-overs in automated driving

Reviewing existing human factors research may reveal what is already known about driver-initiated take-overs and the effects of the driving situation on subjective experience and behavior. To get a better overview of the state of research and potential gaps, relevant studies are chosen and classified. System-initiated take-overs are also considered, and selected experiments are included to be able to draw a comparison between the two types. Selection criteria were the following:

(1) driving studies,

(2) published in peer-reviewed journals or conference proceedings,

(3) between January 2012 and March 2022,

(4) investigated take-overs from the ADS at SAE-level 3 (SAE International, 2021),

(5) in passenger cars.

This resulted in 41 studies. They are classified concerning the earlier presented characteristics: (1) the initiation of take-overs and (2) the criticality of driving situations. In case a publication reports the use of a TOR, the take-over is classified as *system-initiated*, otherwise it is assigned to *driver-initiated*. The time budgets that are reported in the publications are used to rank the driving situations from less to more critical. The classification and ranking of the selected studies are presented in Figure 4. It should be noted that in some experiments, participants experienced two or more different take-over types and various time budgets. Therefore, these studies are listed several times.

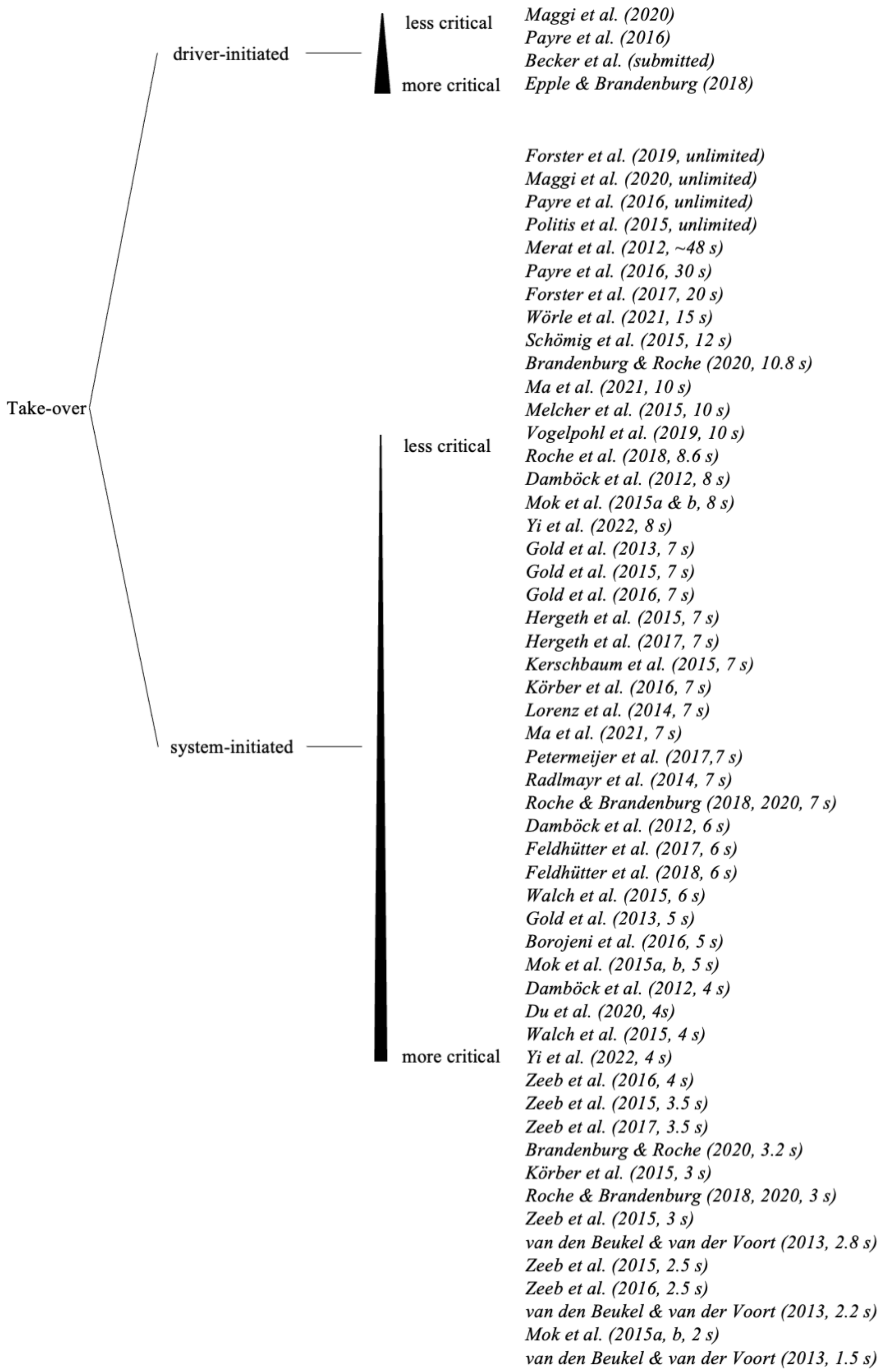

Figure 4: Overview of the characteristics of take-overs with selected driving studies for each take-over type.

Note. The investigated time budgets are presented in parentheses behind each study. Naujoks et al. (2014) did not report the used time budgets, hence, their publication is not listed in this overview. Unlimited time budgets indicate no need to take over to avoid a collision, e.g. because no other vehicles or obstacle were present like in Payre et al. (2016).

1.3.1 Research on driver-initiated take-overs

The presentation of the selected studies of the past ten years in Figure 4 shows that only four focused on driver-initiated take-overs. And this despite the fact that an analysis of seven manufacturers' disengagement reports showed that drivers initiate take-overs with a certain frequency (Lv et al., 2018). The relevance of these disengagements is even more clearly illustrated by an analysis of California's Autonomous Vehicle Tester Program (Boggs et al., 2020). The authors showed that 75 % of all disengagements investigated between September 2014 and November 2018 were driver-initiated, only 25 % system-initiated. Similarly, Epple and Brandenburg (2018) demonstrated in their driving simulator study that the number of driver-initiated take-overs is very high with 85 %. In the remaining 15 % trials, participants did not take over. Even though Payre et al. (2016) exposed their subjects to less critical driving situations with no need to take back control, 8.7 % of them did so. The reports and studies prove that driver-initiated take-overs occur and are relevant events in automated driving. But how do drivers experience such situations and how do they behave when taking back vehicle control?

First of all, when investigating the effects of different situational parameters on subjective experience and take-over behavior, researchers should consider that the parameters cannot be experimentally manipulated as it is possible in system-initiated ones because no TOR exists. The situational parameters can only be varied to evoke a take-over.

Maggi et al. (2020) compared driver-initiated with system-initiated take-overs. For the former ones, participants were instructed to resume control after they completed a non-driving related task (NDRT). Hence, the transition was self-paced and less critical. Driver-initiated take-overs took longer and steering was more extreme than when the take-over was system-initiated. Maggi et al. (2020) assumed that the longer take-over times (TOT) resulted from the drivers' intention to regain situation awareness before taking over when they were the initiators.

Becker et al. (2022) manipulated time budget, vehicle dynamics, and pre-experience trust to evoke intrinsic disengagements in critical driving situations. Participants initiated take-overs more frequently when time budgets were more critical. Vehicle dynamics did not have an effect. About 20 % of driver-initiated take-overs resulted in rear-end collisions (Becker et al., 2022). Neither steering nor brake behavior were reported in this study.

The presentation of the few available studies illustrates the relevance to investigate driver-initiated take-overs and reveals a research gap concerning the effects of such disengagements. To get an idea of how drivers might behave in these situations, studies on system-initiated take-overs are considered.

1.3.2 Research on system-initiated take-overs

A large portion of the human factors research has focused on system-initiated take-overs (see Figure 4) . The addressed research topics are time budget, vehicle dynamics, repeated experience, take-over scenarios, traffic conditions (e.g. traffic), NDRTs, pre-take-over state (e.g. fatigue), and the design of the TOR (e.g. modality, see appendix A for the topics and corresponding studies). In this thesis, I focus on the effects of time budget and vehicle dynamics on subjective criticality and behavior because it is highly likely that the driving situations will vary concerning this aspect.

Research on time budget

Time budgets can be experimentally manipulated in system-initiated take-overs by requesting a take-over when the intended time budget-values are met. A huge portion of researchers investigated the effects of time budgets. These studies showed that higher time budgets are associated with

- higher situation awareness (van den Beukel & van der Voort, 2013),
- lower criticality ratings (Roche & Brandenburg, 2018, 2020; Yi et al., 2022),
- higher comfort ratings (Mok, Johns, Lee, Ive, et al., 2015),
- higher trustworthiness ratings (Mok, Johns, Lee, Miller, et al., 2015),
- higher TOTs (Roche & Brandenburg, 2018, 2020),
- lower decelerations (Roche & Brandenburg, 2018, 2020),
- better lateral vehicle control (Mok, Johns, Lee, Ive, et al., 2015; Mok, Johns, Lee, Miller, et al., 2015),
- smaller steering wheel angles (Roche & Brandenburg, 2018, 2020),
- fewer road departures (Mok, Johns, Lee, Ive, et al., 2015; Mok, Johns, Lee, Miller, et al., 2015),
- and fewer driving errors (Damböck et al., 2012; van den Beukel & van der Voort, 2013).

In contrast to these studies, Walch et al. (2015) found neither subjective nor behavioral differences between the different time budgets. This may be because they compared situations with 4 and 6 s time budgets, while others investigated larger ranges, e.g. 3 and 7 s (Roche & Brandenburg, 2018, 2020). Additionally, Walch et al. (2015) analyzed TOTs and brake behavior while others found significant differences in steering behavior and driving errors, e.g. Mok, Johns, Lee, Ive, et al. (2015).

Next to the range, the magnitudes of the investigated time budgets differed between the studies. The magnitude spans from 0.5 s (van den Beukel & van der Voort, 2013) to unlimited (Maggi et al., 2020; Payre et al., 2016). Unlimited time budgets indicate no need to take over, e.g. because no other vehicles or obstacles were present. Van den Beukel and van der Voort (2013) and Mok, Johns, Lee, Miller, et al. (2015) realized the lowest time budgets of the presented studies. The former investigated the effects of time-critical brake situations on collision frequency and situation awareness. They used THWs of 0.5 s, 1.0 s, and 1.5 s (corresponding to TTCs of 1.5 s, 2.2 s, and 2.8 s) to a strongly braking vehicle (-8 m/s^2). The results indicated that more time for a take-over leads to more successful take-overs and higher situation awareness (van den Beukel & van der Voort, 2013). Mok, Johns, Lee, Miller, et al. (2015) also investigated take-over situations with low time budgets: 2, 5, or 8 s TTC to a construction site requiring a lane change. With lower TTCs, steering behavior varied more, more lane departures occurred, and lower trustworthiness ratings were recorded (Mok, Johns, Lee, Miller, et al., 2015).

These studies demonstrate that larger time budgets affect subjective experience and improve take-over behavior. The likelihood to detect effects may depend on the range and magnitude of the investigated time budgets. However, it is unknown whether drivers react similarly when they initiate the take-over themselves.

Research on vehicle dynamics

Like time budget, vehicle dynamics can be experimentally manipulated in system-initiated take-overs. However, only one study is available employing vehicle dynamics as an independent variable. In a driving simulator, Hu et al. (2019) investigated the passengers' hazard perception by varying velocity, deceleration, and distance between the ego-vehicle and an obstacle in a braking maneuver. Decelerations correspond to the vehicle dynamics and were varied on three levels: -1 m/s^2, -3 m/s^2, and -6 m/s^2. The participants were passengers and rated the hazard perception after experiencing the situation. Hu et al. (2019) reported that velocity, deceleration, and distance affected the experience of the driving situations. It should be noted that the authors investigated the experience of passengers, not the experience of drivers. Hence, take-over behavior was not analyzed. Other studies used vehicle dynamics or similar parameters as dependent variables, e.g. Du et al. (2020) and Gold et al. (2013).

Driving studies on the effects of vehicle dynamics on subjective experience (from the drivers' perspective) and take-over behavior are missing. Hence, it is questionable how drivers experience dynamic driving situations and how they behave when taking over.

1.3.3 Conclusion

The overview over the current state of research showed that driver-initiated take-overs are mostly neglected while most studies focused on system-initiated ones. Research on the latter ones demonstrated that low time budgets lead to extreme take-over behavior and may have serious consequences. Only one study is available that deals with vehicle dynamics. However, these results cannot simply be transferred to driver-initiated take-overs. Hence, the effects of situational criticality on perception and behavior of these take-overs are unknown.

1.4 Research on repeatedly taking over vehicle control

In addition to driver-initiated take-overs, investigating the effects of repeatedly taking over vehicle control is of interest because most driving simulator studies test several trials in one session with a much higher frequency than in realistic traffic conditions. Familiarity with the driving simulator, training effects concerning the take-over process, adopted trust in the ADS, or fatigue may set in (Feldhütter et al., 2018; Forster et al., 2019; Hergeth et al., 2017; Roche et al., 2018). Consequently, take-over behavior may adapt. In this case, changes in both directions are possible. On the one hand, the reactions might improve due to increasing familiarity with the simulator or training of the take-over process. Participants may learn to compensate for flaws of the test setting or how to react to specific cues. On the other hand, the behavior might deteriorate due to increasing fatigue or trust in the system. Hence, observations of the umpteenth trial might not be as robust as those of the first trials. This will affect the validity of the results if the change is disregarded.

Some of the selected studies compare the different trials with each other, hence, a first insight into the effects of the repeated experience of take-over trials is possible. Participants in Hergeth et al. (2017) experienced two take-overs and rated the criticality afterward. The ratings decreased, even though, participants experienced system limits (Hergeth et al., 2017). Besides, TOTs decreased, minimal TTCs increased, and resulting maximal accelerations decreased from the first to the second take-over. The authors argued that the effects are due to the increasing familiarity. In line with Hergeth et al. (2017), Payre et al. (2016) observed decreasing TOTs from the first to the second take-over. In contrast, Brandenburg and Roche (2020) and Roche et al. (2018) observed different effects. They investigated six take-overs. A change of TOTs was not observed (Brandenburg & Roche, 2020; Roche et al., 2018), but deceleration increased from the first to third experience (Brandenburg & Roche, 2020). Forster et al. (2019) investigated several transitions to higher and lower levels of automated driving. They observed a decrease of TOTs, experimenter ratings, and error rates within the first trials and stabilization after three trials for transitions between different levels of automation. When deactivating the driving automation, no behavioral change was observed.

The studies on the repeated experience of take-overs show that criticality ratings and behavior change to a certain extent over trials. The trend is sometimes positive, sometimes negative, sometimes non-existent. However, most of the studies investigated the effects of two or three trials, the changes over more take-overs were barely researched. Increasing training or familiarity may be a reason for a change. Besides, the evolution of fatigue was not assessed, even though Feldhütter et al. (2018) showed that objective and subjective measures of sleepiness, i.e. expert rating and eye-tracking, increase when driving automated for a longer period. Increasing fatigue can in turn impair the take-over quality. Examining more trials may reveal the long-term changes of behavior and indicate whether the results of driving studies with repeated measures are valid.

1.5 Criticality assessment of driving situations

In the present thesis, not only the criticality of driving situations but also its assessment plays a leading role. It would indicate whether the variations of situational parameters such as time budget and vehicle dynamics are perceived and whether they interact. Assessing the criticality may support the interpretation of observed take-over behavior. Additionally, different driving situations might be compared concerning this dimension.

The perceived criticality of driving situations can be assessed by self-reported measures, e.g. scales. They are easy to handle and their application is economical because no further equipment is needed than paper and pencil or tablet (Moosbrugger & Kelava, 2020). Also, the data processing is parsimonious because one value is available per event (Moosbrugger & Kelava, 2020).

Various scales with different response formats have been used in the past. They can be summarized into three types: (1) rating, (2) two-step rating, and (3) visual analog scales. All of them are presented briefly. First, different ratings scales were used by Feldhütter et al. (2018), Radlmayr et al. (2018), and Roche and Brandenburg (2018, 2020). Feldhütter et al. (2018) assessed the criticality of take-over situations with a 10-points rating scale ranging from ‘not critical’ to ‘extremely critical’. Radlmayr et al. (2018) used a 7-point Likert scale, but they did not state whether they labeled the points. Similarly, Roche and Brandenburg (2018, 2020) surveyed perceived criticality using an 8-point Likert scale ranging from ‚not critical’ (0 pts.) to ‚very critical‘ (7 pts.). They visualized the increasing criticality from left to right with a triangle (see Figure 5). The wording of the question is available. This scale was already used in three driving simulator studies of the research group (Brandenburg & Roche, 2020; Roche & Brandenburg, 2018, 2020).

Second, in Winkler et al. (2018), participants rated the criticality of different driving situations on a two-step rating scale based on Heller (1982). First, one of five verbal categories had to be selected ranging from ‘very low’ to ‘very high’. Second, to specify the evaluation, respondents had to choose one of three subcategories ‘-‘, ’0’, or ‘+’ (see Figure 6). This resulted in a 15-points scale (Heller, 1982). Similar to Heller (1982), the *Scale of Criticality Assessment of driving situations* (SCA, see Figure 7) is based on a two-step rating procedure. It is a modified version of the *Scale for the Assessment of the Experienced Degree of Disturbance* introduced by Neukum and Krüger (2003). First, respondents choose one of the verbal categories ‘imperceptible’, ‘harmless’, ‘unpleasant’, ‘dangerous’, and ‘uncontrollable’ (translation based on Naujoks et al., 2017). Second, they are requested to specify their answer by selecting one of the numerical subcategories, resulting in an eleven-point rating scale.

Third, Banet and Bellet (2008) used a visual analog scale ranging from ‘not critical’ to ‘high level of criticality’. Evaluations were provided by moving and setting the slider to a value between 0 to 100 pts. This scale allows for a more granular rating.

Disadvantages of the scales used by Banet and Bellet (2008), Feldhütter et al. (2018), and Radlmayr et al. (2018) are that the authors provided neither a visualization of the scales in their publications nor the wordings of the questions. Hence, these scales are not suited for further investigation. For the rating scales, visualization and wording are only available for the one introduced by Roche and Brandenburg (2018, 2020). It reflects a certain granularity of the criticality rating and establishes an equidistance between the scale points by avoiding additional verbal labeling. Due to that, it is included in the studies of the present thesis.

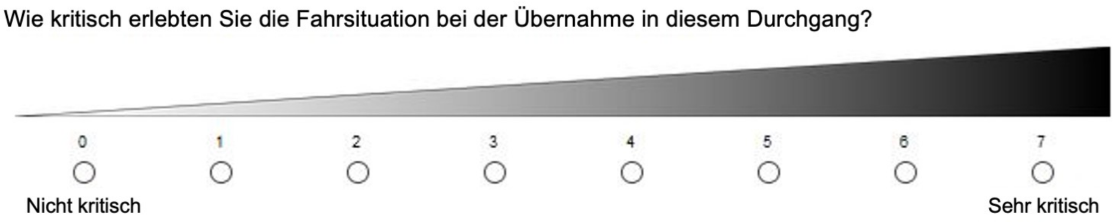

Figure 5: Criticality scale by Roche and Brandenburg (2020, 2018) in German.

Note. Translation of the question into English: ‘How critical did you perceive the driving situation during the take-over in this trial?’. The poles are labeled ‘not critical’ and ‘very critical’.

very low			rather low			moderate			rather high			very high		
-	0	+	-	0	+	-	0	+	-	0	+	-	0	+

Figure 6: Two-step rating scale used by Winkler et al. (2018, p. 4), based on Heller (1982).

Figure 7: Scale of Criticality Assessment (SCA, Neukum et al., 2008, p. 4).

Note. The translation of the verbal categories into English is ‘imperceptible’, ‘harmless’, ‘unpleasant’, ‘dangerous’, and ‘uncontrollable’ (translation based on Naujoks et al., 2017).

The SCA developed by Neukum et al. (2008) was already used in several driving studies (Hergeth et al., 2017; Naujoks et al., 2017; Naujoks & Neukum, 2014; Neukum et al., 2008; Purucker et al., 2014; Siebert et al., 2014; Tscharn et al., 2018; Wörle et al., 2021; Yi et al., 2022). Furthermore, its visualization is available, and the question of the *original* scale and an extensive description of the verbal categories are reported in Neukum and Krüger (2003). However, neither an adjusted question nor description is available for the *modified* version on criticality. Additionally, the equidistance between numerical values is questionable due to the verbal categories. For example, the perceived difference of two ratings from one verbal category (e.g. 'harmless' 2 and 3 pts.) might be smaller than the distance of two evaluations between two categories (e.g. 'harmless' 3 and 'unpleasant' 4 pts.), even though the numerical differences are the same. Besides, internal pre-tests showed that instructing the scale and analyzing the results are time-consuming. And still, the rating is error-prone since participants partially do not understand the correct usage. Hence, verbal and numerical ratings given by respondents do not correspond. Finally, even though Neukum et al. (2008) stated that the SCA was validated in internal studies, no publications on the validation are available. Because this scale was used more often than the one presented by Winkler et al. (2018), it is selected for further investigation.

To conclude, various scales have been employed to assess the criticality of driving situations. However, most of the available publications entail several disadvantages, such as missing visualization. Additionally, the scales lack validation. Hence, uncertainty persists which scale is appropriate to assess the criticality of driving situations. Modifying existing scales and validating them is necessary to overcome this gap. The scales developed by Roche and Brandenburg (2018, 2020) and the one introduced by Neukum et al. (2008) were selected to be validated.

1.6 Research question

The presentation of current research showed that driver-initiated take-overs were barely investigated in the past. Hence, it is unclear how drivers behave when initiating a take-over in critical and dynamic driving situations. This leads to the main research question (RQ) of the thesis:

What might happen if drivers take over control in highly time-critical and dynamic driving situations?

The aspects presented in the introduction were investigated in three driving simulator studies which form the basis of this thesis. Study 1 was planned in cooperation with Oliver

Blum. He also performed the data collection. In this experiment, the two criticality rating scales were validated and the effects of repeatedly taking over control were investigated. Studies 2 and 3 were designed and conducted in the course of the interdisciplinary research project 'Analysis and Support of Driver Interventions in Dynamic: Critical Situations during Highly Automated Driving' funded by the German Research Community (Deutsche Forschungsgemeinschaft, DFG, Grant No. 326727090). The two studies focused on the effects of driver-initiated take-overs in critical driving situations on subjective criticality and driver behavior. Jun.-Prof. Dr. Stefan Brandenburg, M. Sc. Sandra Becker, Prof. Dr.-Ing. Steffen Müller, M. Sc. Thang Nguyen, and Prof. Dr. Manfred Thüring supported the design and preparation of the two experiments. Marc Buchholz, Jan Haentjes, and Philipp Wittke assisted with data collection.

2 Empirical studies

The three studies that were conducted to answer the research question are briefly presented in the following. An overview of the included publications can be found in appendix B. Details about the methods, materials, exact results, and discussion can be found in the original publications in appendices C, D, and E.

2.1 Study 1: Assessing the criticality of driving situations after automated driving: Validation of two scales

2.1.1 Introduction

An assessment of the criticality of driving situations is essential to reveal whether its variations are perceived and how different situational parameters such as time budget and vehicle dynamics interact. Scales may be used to measure criticality. To ensure that they are truly measuring what they are supposed to measure (Hartig et al., 2008), they need to be validated. According to Frey (2018), validity is the most important quality criterion, even more important than reliability and objectivity. It increases the generalizability of the results. Hence, a prerequisite to answer a part of the research question was a validated scale.

As section 1.5 showed, several of the scales were used in previous driving studies, but they entail some disadvantages, such as missing visualization. Additionally, they lack validation. In study 1, two suitable rating scales were validated. The manuscript is published in the journal *Accident Analysis and Prevention*, Volume 159 (Roche, 2021, see appendix C).

The scale developed in 2018 by Roche and Brandenburg (see Figure 5) was revised to address the discussed disadvantages of existing scales (for more details see 2.1.2 or appendix C). This revision resulted in the Criticality Rating Scale (CRS, see Figure 8). The validity of the CRS and the SCA (see Figure 7) was surveyed. Besides, the validities of both scales were compared. This leads to the following three research questions of study 1:

RQ 1: Is the Criticality Rating Scale (CRS) a valid tool for the assessment of the criticality of driving situations?

RQ 2: Is the Scale of Criticality Assessment of driving situations (SCA) a valid tool for the assessment of the criticality of driving situations?

RQ 3: Do both scales differ regarding their validity?

Next to the validation, another aspect was considered in study 1. The state of research showed that criticality ratings and behavior tend to change when drivers take over repeatedly. Mostly, effects were investigated over two or three trials. In some studies, behavior improved, while it did not change or deteriorated in others. One study showed that criticality ratings decreased and take-over behavior improved from the first to the second trial (Hergeth et al., 2017). Examining the effects of experiencing more than three take-overs will reveal whether the perceived criticality and behavior changes in the long term. It may enable an estimation of the validity of the obtained results from driving studies testing several take-overs in a row. These aspects were addressed in the fourth research question. Additionally, when experiencing more trials in a row, fatigue may increase leading to a deterioration of take-over behavior. Assessing the development of sleepiness may help interpret results. Therefore, it was also considered in the fourth research question. It was operationalized by ratings of effort because de Waard (2002) stated that progressing fatigue is reflected in increasing effort that is invested to cope with the situation:

RQ 4: Do drivers' criticality and effort ratings and take-over behavior change over the repeated experience of take-overs?

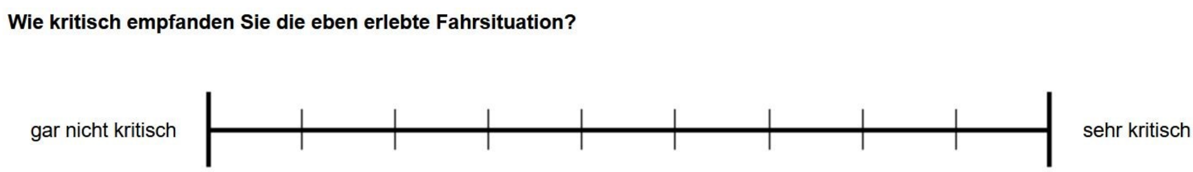

Figure 8: Criticality Rating Scale (CRS), a revised version of the scale used by Roche and Brandenburg (2018, 2020).

2.1.2 Method

First, the already existing scale developed in-house (see Figure 5) was modified. The complexity of the question was reduced, as recommended by Moosbrugger and Kevala (2020). A continuous scale with ten levels was used to collect the responses. The design was inspired by the NASA-TLX which is a very frequently used and validated tool for the assessment of subjective workload (Hart, 2006; Hart & Staveland, 1988). In contrast to the NASA-TLX, a middle answer category was avoided because it was observed that it is commonly chosen as an alternative when respondents do not understand the question or refuse to answer (Moosbrugger & Kelava, 2020). Besides, this measure prevented errors of central tendency (Moosbrugger & Kelava, 2020). The verbal labeling of the scale poles was based on empirical results from Rohrmann (1978) demonstrating that the terms ‘gar nicht’ (Engl. ‘not at all’) and ‘sehr’ (Engl. ‘very’) were consistently interpreted as extrema by the participants. To allow an equidistant gradation of the response format, any further naming or numerical anchors were omitted. The scale is expected to be rapid in application due to its short instruction. The revision resulted in the CRS presented in Figure 8. More details concerning the scale construction can be found in the corresponding publication (see appendix C).

The research questions were investigated in a static driving simulator study of the Department of Psychology and Ergonomics of Technische Universität Berlin. Twenty-five participants (13 women, 12 men) completed the experiment. After a familiarization and training phase, they experienced two blocks of five trials, ten in total. The two blocks differed in respect to the scale that was used to provide the rating: the SCA or CRS. The sequence was balanced across participants. In each trial, a take-over was necessary due to an obstacle on the participants’ lane (see Figure 9). As discussed in the section on time budget, TTC is a suitable measure to describe the criticality of driving situations with a stationary obstacle. Hence, TTC at the moment of the TOR was used to vary the criticality. Five equidistant TTC-values were realized: 2.5, 3.0, 3.5, 4.0, and 4.5 s. Over all participants, each TTC was experienced at every position five times, e.g. TTC 2.5 s was presented first for five participants. Hence, the experimental design was a 2 (rating scale) x 5 (TTC-value)-within-subjects design.

In each trial, participants followed a lead vehicle in automated driving mode. Participants had to take over vehicle control upon request and change lanes to avoid a collision with the obstacle (see Figure 9). With a valid scale, a variation of the criticality of the driving situations should be reflected in the corresponding ratings.

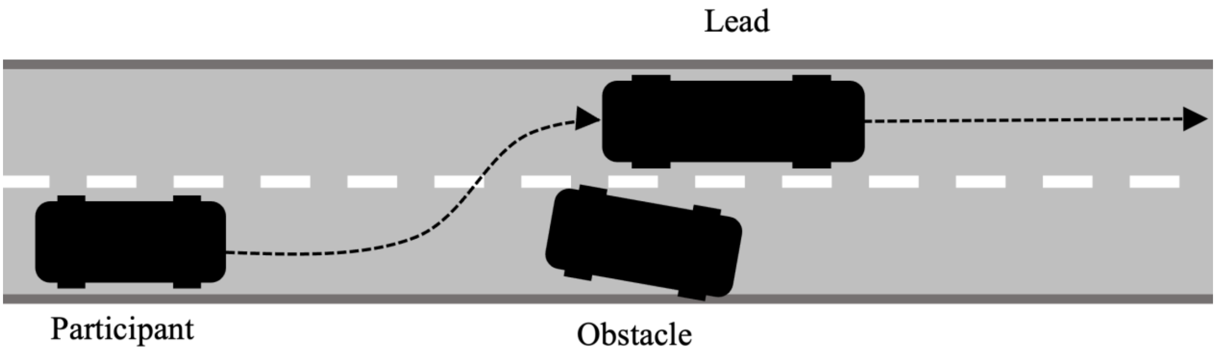

Figure 9: Schematic illustration of the lane change situation.
Note. The participants' vehicle follows the lead vehicle. It changes lanes due to the obstacle in its lane.

TTCs and criticality ratings for each scale were correlated and evaluated (RQ 1 and 2). For the comparison between the SCA and CRS, it was tested whether the correlation coefficients of both scales differed significantly (RQ 3). This approach was based on a method suggested by Eid et al. (2017). For details, see appendix C, section 2.5. Using this method, the convergent validity of the scales is investigated. Driving data for each take-over were recorded to investigate the effects of repeated experience on behavior (RQ 4). Participants provided their ratings of perceived effort after each trial with the subscale of the NASA TLX (Hart & Staveland, 1988). Linear mixed-effects models were used to evaluate these effects.

2.1.3 Results and discussion

The ratings of both scales correlated strongly with the TTC-values. This indicated that the SCA and CRS are valid for the assessment of the criticality of take-over situations (RQ 1 and 2). However, convergent validity was tested only, while other types were not considered. The reasons were that this type of validity is an important property of assessment tools and in this case, it was easy to investigate because ratings could be correlated with TTC-values. Future studies should test further validities such as discriminant validity. Besides, the scales were validated in lane change situations in which criticality was varied by TTC. Hence, a generalization to other driving scenarios and other variations of criticalities is debatable.

The comparison revealed that neither the SCA nor the CRS is superior to the other; they are equally well suited to assess criticality (RQ 3). This is noteworthy because the two scales use different scale designs and vary concerning the instructions. Internal pre-tests and this experiment showed that instructing the SCA is more time-consuming. Besides, its application is more error-prone than the CRS, as sometimes the markings of the first and second rating step do not correspond, e.g. a participant marks the verbal category 'harmless' and the numerical subcategory 5. Hence, usage of the CRS is recommended.

Repeatedly taking over did neither affect criticality nor perceived effort ratings nor take-over behavior. Except maximal brake pedal position decreased over trials (RQ 4).

In retrospect, two aspects of this study should have been done differently. First, perceived effort was assessed as an indicator of fatigue. However, another tool, e.g. the Karolinska Sleepiness Scale (Shahid et al., 2011), would have been a more appropriate and direct way to measure fatigue. Therefore, I would use this scale in future studies. Second, the change in the criticality ratings over trials was recorded using a scale whose reliability has not yet been checked. Hence, the results of the criticality ratings concerning repeated take-overs should be viewed with caution and are not considered further in this thesis.

2.1.4 Conclusion

Based on the results of study 1, it can be concluded that both scales are appropriate to assess the criticality of take-over situations. The CRS was selected to be applied in studies 2 and 3 because it was observed that it holds a higher test efficiency and lower susceptibility to errors than the SCA.

Study 1 indicated that neither effort ratings nor take-over behavior of driving simulator studies are affected by frequent repetitions of take-over situations.

2.2 Study 2: What happens when drivers of highly automated vehicles take over control in critical brake situations?

2.2.1 Introduction

Even though automated driving systems (ADS) are designed to relieve the drivers, driver-initiated take-overs may still occur. This can happen any time because the option to override the ADS must always be available, also in critical and dynamic driving situations (United Nations Economic Commission for Europe, 2014).

The presentation of the research demonstrated that driver-initiated take-overs were hardly examined, although different studies showed that drivers do initiate take-overs. These control shifts in critical driving situations are of particular interest, as appropriate behavior can be crucial here to avert worse. Hence, the aim of study 2 was to investigate how drivers perceive time-critical and dynamic brake situations, and how they behave and perform when initiating a take-over. The experiment is published in the journal *Accident Analysis and Prevention*, Volume 144 (Roche et al., 2020). It can be found in appendix D.

The following three research questions were investigated in study 2:

RQ 1: Does the criticality of the driving situations impact the criticality rating indicating that drivers can discriminate between the different degrees of criticality?

RQ 2: Does the criticality of the driving situations influence take-over behavior?

RQ 3: Do drivers deliver an appropriate performance when they take over?

RQ 2 refers to take-over behavior. In this thesis, it is understood as the pure observation of the action without evaluating it. Dependent variables that characterized the behavior are TOTs and maximal decelerations. RQ 3 deals with the take-over performance which is here defined as the evaluation and the consequences of the actions. Dependent variables are lane departures, collisions, and comparisons of deceleration behavior to the actions of the automation. For the comparison, the maximal deceleration recorded per participant and trial was subtracted from the data of the automation for the corresponding trial. The comparison was used to determine whether the extent of decelerations was adequate to avoid collisions with the merging vehicle. A significantly stronger reaction in comparison to the ADS was interpreted as an overreaction. This method is a rather novel and unusual approach to evaluate take-over performance.

2.2.2 Method

The three research questions were addressed in a study conducted in the driving simulator of the Department of Automotive Engineering, Technische Universität Berlin. Forty-two subjects (19 women, 23 men) took part. After a training session, the participants experienced nine experimental and six filler trials, 15 in total. In each one, they followed a lead vehicle in automated driving mode for about 1.5 min. Then, in the experimental trials, another car merged into the lane between the participants' and the lead vehicle and braked strongly. An exemplary presentation of the maneuver is visualized in Figure 10. The criticality of these brake situations was varied by time budget in terms of time headway (THW) and by vehicle dynamics, realized by traction usage (TU). THW was defined as the distance between the participants' and the merging vehicle; three equidistant THW-levels were implemented: 0.34, 0.21, and 0.08 s. TU was defined by the ego vehicle's deceleration of the advanced emergency braking system, three levels were realized: 0.44, 0.64, and 0.84. This resulted in a 3 (THW) x 3 (TU) within-subjects-design. Details on the randomization can be found in the corresponding publication (see appendix D). Participants were triggered to take over control by an acoustic cue when the intended combinations of THW and TU were met. The ADS was designed to manage the situation safely by decelerating appropriately in case the drivers did not take back control. In the filler trials, participants were either triggered to take over without a merging vehicle, or no trigger was presented.

Criticality ratings were collected after each trial with the CRS (Roche, 2021) and behavior was recorded. The effects of THW and TU on criticality ratings, take-over behavior, and deceleration difference to the automation were analyzed with linear mixed-effects models (RQ 1, 2, and 3). The frequencies of collisions and lane changes were analyzed with logistic models (RQ 3).

2.2.3 Results and discussion

The analysis showed that criticality ratings and maximal decelerations increased significantly when THW or TU became more critical (RQ 1 and 2). No interaction effect on criticality ratings was observed. The interaction effect on maximal decelerations showed that the effect of TU was stronger when THW was less critical. In contrast to the expectations, TOTs were neither affected by THW nor by TU (RQ 2). A reason for this result could be that the THW-values were too low to cause a difference. Another one might be that the take-overs by the participants were very fast. Hence, THW or TU could barely affect them because of a floor effect.

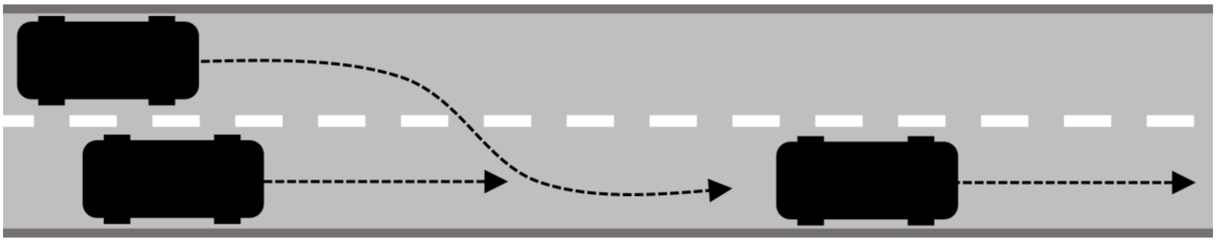

Figure 10: Schematic illustration of the brake situation.
Note. The participants' automated vehicle follows the lead vehicle with a constant velocity and distance. The merging vehicle decelerates as soon as it has merged into the participants' lane.

The results revealed that participants decelerated significantly more than the automation, hence, unnecessarily strong (RQ 3). This difference was larger when TU was less critical. These findings can be interpreted as an overreaction since small deceleration would have been sufficient to establish a safe distance to the merging vehicle. Besides, lane changes occurred, even though they were not necessary to avoid a collision. Both findings indicate a hazardous behavior when taking over because strong decelerations and lane changes may destabilize the vehicle or result in collisions with following vehicles or on the adjacent lanes (RQ 3). A collision with the merging vehicle occurred in only three out of 357 trials. Due to the small number, no statistical analysis was performed on collision frequency.

2.2.4 Conclusion

To conclude, more critical brake situations in terms of low THWs and high TUs lead to higher criticality ratings and higher maximal decelerations. The variable 'deviation of the maximal deceleration from the automation' revealed that the overreaction was higher when TU was less critical. However, the variable was not suitable to derive in which conditions participants decelerated stronger than necessary.

Despite these results, it is questionable whether drivers would behave and perform similarly when taking over in different driving situations. Investigating further scenarios would extend the reported findings. This leads to study 3, examining take-overs in lane change situations.

2.3 Study 3: What Happens when Drivers of Automated Vehicles Take Over Control in Critical Lane Change Situations?

2.3.1 Introduction

Study 2 showed that the situational parameters time budget and vehicle dynamics affect the subjective criticality, the behavior, and the performance when drivers take over in critical brake situations and that take-overs in critical and dynamic brake situations are dangerous. However, also the driving conditions themselves may influence the dependent variables. Therefore, it is questionable whether similar results would be obtained when investigating different driving situations.

Lane changes are suitable for such an investigation because it is a typical maneuver and can be critical and dynamic. Due to that, study 3 focuses on lane change situations. In line with the previous experiment, it investigates how drivers perceive time-critical and dynamic lane changes, how they behave, and whether they react appropriately when initiating a take-over. This study is published in the journal *Transportation Research Part F: Psychology and Behavior* (Roche et al., 2022, see appendix E).

In study 3, the same research questions were investigated as in study 2:

RQ 1: Does the criticality of driving situations impact the criticality ratings indicating that drivers can discriminate between the different degrees of criticality?

RQ 2: Does the criticality of the driving situations influence take-over behavior?

RQ 3: Do drivers deliver an appropriate performance when they take over?

Like in study 2, RQ 2 refers to take-over behavior, while RQ 3 deals with the performance. Here, TOTs, maximal steering wheel angles, and maximal deceleration were used to answer RQ 2. To evaluate take-over performance, lane departures, collisions, and comparisons of deceleration and steering behavior to the automation were analyzed. Again, significant deviations from the ADS were interpreted as overreaction.

2.3.2 Method

Study 3 was conducted in the same driving simulator as study 2. After a training session, sixty participants (28 women, 32 men) experienced 14 double lane change situations, eight experimental and six filler trials. Each run started with an automated phase. The participants' car followed a lead vehicle for about 1 min. Then, the preceding car performed a double lane change because of obstacles in the lanes. The maneuver is schematically presented in 11. In the experimental trials, the take-over was either triggered before the first or the second lane change. The criticality of the driving situations was varied by time budget, here time-to-collision (TTC), and traction usage (TU). TTC was defined by the distance to the obstacle at the moment of the trigger. Two TTC-levels were implemented: 2.1 s and 1.2 s. TU was varied by the lateral acceleration due to the lane change maneuver performed by the ADS. Two TU-levels were realized: 0.24 and 0.38. The take-overs were provoked by an acoustic cue with the intention that participants take back vehicle control in comparable driving situations. In the filler trials, no trigger was presented during the double lane changes. The ADS was designed to manage the situation safely by steering around the obstacles and following the lead vehicle on the same trajectory in case the drivers did not take over.

The two lane changes were analyzed separately because they differed in many respects, e.g. predictability of the take-overs and the event preceding the trigger. This resulted in two 2 (TTC) x 2 (TU) within-subjects-designs, one for the first lane change and one for the second. Criticality ratings were collected after each trial with the CRS validated in study 1 (Roche, 2021). TOTs, maximal steering wheel angles, maximal decelerations, lane departures, and collisions were recorded for each trial. The metric variables were examined with linear mixed-effects models (RQ 1 and 2). The binary ones were analyzed with logistic models (RQ 3). In contrast to study 2, not the difference to the automation was calculated, but t-tests for each condition were used to check whether participants' steering and decelerating deviated from the automation (RQ 3). This method reveals differences for each condition rather than investigating how much time budget and vehicle dynamics contribute to the difference.

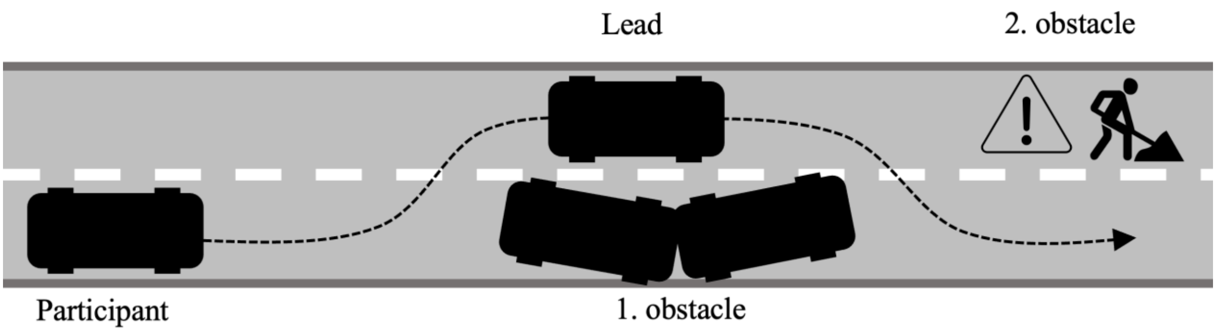

Figure 11: Schematic illustration of the double lane change.

Note. "Warning sign" by Alfa Design from Noun Project. "Work in progress sign" by Gleb Khorunzhiy from Noun Project. Both licensed under CC BY 4.0

Note. The participants' automated vehicle follows the lead vehicle on the same trajectory. The trigger was either presented before the first or the second obstacle.

2.3.3 Results and discussion

The analysis showed that lower TTC and higher TU led to increased criticality ratings. Only in the first lane change, TTC did not have an effect. On the one hand, this may indicate that participants truly evaluated the driving situations and, thus, can discriminate between different degrees of TTC and TU (RQ 1). On the other hand, it could also be that their answers were influenced by the outcome of their action, e.g. collision or lane departure, since they provided their rating after each trial. An additional analysis showed that the TOTs correlated with the ratings. This indicates that critical situations were perceived as such, and participants reacted accordingly. However, the ratings also correlated with collision frequency, suggesting that an a-posteriori assessment may be biased by the outcome of the maneuver to some extent.

With more critical TTC and TU, lower take-over times (only in the second lane change), higher maximal decelerations, and higher maximal steering wheel angles were observed (RQ 2). In the first lane change, more collisions occurred when TTC was less critical and TU was more critical. In both maneuver parts, participants departed from the lane more often with more critical TTC. Participants steered more than necessary in most conditions in the first lane change and in one condition in the second lane change, especially when TTC was more critical. This might be the reason for the higher lane departure probability with more critical TTCs in both lane changes. Additionally, participants decelerated more than the automation under all conditions. In total, the results on collisions, lane departures, steering, and deceleration indicate a poor and hazardous take-over performance (RQ 3). Especially, the observed lane departures and collisions deserve attention because they may lead to serious damages and would not have happened if the driver had not taken over vehicle control.

The observed interactions of criticality ratings, TOTs, maximal steering wheel angles, and lane departures in the second lane changes showed that the effect of TTC was stronger when TU was less critical. It could be that more critical TU was perceived as a threat to the drivers' safety and, thus, captured attention. Hence, fewer cognitive resources were available for processing other stimuli such as TTC (Kahneman, 1973). When TU was less critical, more cognitive resources were free to focus on TTC. Hence, TTC could have a stronger effect on the dependent variables.

2.3.4 Conclusion

To conclude, the results of study 3 demonstrated that more critical lane change situations concerning TTC and TU are perceived as such – at least to a certain extent. Take-overs in these situations lead to more extreme steering and decelerations. In most conditions, drivers tend to overreact by steering and deceleration and worsen the situation, i.e. they cause lane departures and collisions.

3 Discussion

Previous research intensively investigated system-initiated take-overs. However, driver-initiated take-overs have barely been studied. The effects of vehicle dynamics on take-over behavior were not considered at all. Hence, the research question of the present thesis was what might happen if drivers initiate take-overs in highly time-critical and dynamic driving situations.

To address the research question, three driving simulator studies were conducted. In study 1, two rating scales were validated to solidly survey the subjective criticality of the take-overs in the following studies. In addition, the change of effort ratings and behavior over the repeated experience was examined. In studies 2 and 3, subjective criticality, take-over behavior, and performance were investigated in brake (study 2) and lane change situations (study 3) of varying criticality.

A difference between the three studies should be kept in mind when discussing and comparing them. In the first and third experiment, lane change maneuvers were investigated. Here, TTC was varied because it is usually employed to describe the criticality of such situations (see the section on Time budget). Hence, the reported values in these two studies are comparable. In the second experiment, brake situations were examined. The time budgets referred to moving objects, the merging and braking vehicles. Hence, THW was used to vary and describe the time budgets. Besides, TU was not considered in study 1, while it was altered in longitudinal direction in study 2 and lateral direction in study 3. Thus, the absolute time budget- and TU-values of the lane change and the brake maneuvers represent a different criticality. Hence, the three studies should be compared with caution.

3.1 Discussion of the results

The discussion of the results follows the three main aspects of the present thesis: (1) criticality assessment, (2) repeatedly taking over, and (3) effects of time budget and vehicle dynamics in driver-initiated take-overs.

3.1.1 Criticality assessment of driving situations

It was demonstrated that the CRS and the SCA are equally valid scales to assess the criticality of driving situations. This is an important finding because it ensures the quality of the ratings obtained with the CRS and the SCA. However, validation can only ever be as good as the method used. Limitations of the study design are presented in the corresponding publication (appendix C) and section 3.3. They discuss that there is no doubt that the scales need to be validated in further driving situations and with different variations of criticality. Also, testing the discriminant validity, reliability, and objectivity of the scales would increase the generalizability and value of the results. Besides, validation is not the only requirement to ensure that the ratings correspond with the perception. Scales also entail several disadvantages which should be considered.

First, answering a question requires cognitive effort to understand it, recall the relevant information from memory, evaluate its relevance, and select an appropriate answer (Bogner & Landrock, 2016). Depending on the difficulty of the question, the cognitive ability of the respondents, and their motivation to reply truthfully, the answers can be biased (Bogner & Landrock, 2016).

Second, further response biases of scales are possible. These are the tendencies for extreme or central responding or interviewer effects (Bogner & Landrock, 2016; Furnham, 1986). Unambiguous and simple wording of the question, closed response format, and no middle answer category are methods that may counteract the mentioned biases (Bogner & Landrock, 2016; Moosbrugger & Kelava, 2020). These options to reduce distortion were considered in the development of the CRS. Additionally, CRS- and SCA-ratings were provided in absence of the interviewer in the presented studies. Hence, it can be assumed that these response biases were reduced to a certain extent.

Third, scales do not allow a continuous measurement over time. Measuring perceived criticality permanently would have been an interesting aspect in the investigated driving situations because the time budgets and vehicle dynamics changed constantly and were not static throughout one trial. The reported values of the independent variables only represent the criticality of one selected moment. Hence, the collected ratings may refer to different manifestations of time budgets and vehicle dynamics than the intended ones. Psychophysiological parameters might be suitable for continuous measurement. However, they also still need to be validated as indicators for situational criticality. Yi et al. (2022) took a first step in that direction. They showed that more critical take-over situations led to higher skin conductance levels, higher heart rates, and larger pupil diameter. The psychophysiological variables correlated with the criticality ratings provided on the SCA (Yi et al., 2022). Future examinations should pick up here.

Forth, without a continuous measurement, the point of time of the data collection has to be determined. The additional analysis in study 3 showed that a-posterior criticality ratings significantly correlated with TOTs and frequency of collisions (Roche et al., 2022). This means that the ratings corresponded to the behavior but that the outcome was also considered when evaluating the driving situation after experiencing it. On the one hand, a bias of the ratings by the take-over behavior and the outcome of the maneuver cannot be excluded. On the other hand, a real-time evaluation may be biased by time pressure.

Fifth, individual parameters may also affect the perception of driving situations and, thus, bias the evaluations. For example, in Banet and Bellet (2008), motorcyclists and car drivers watched video sequences of driving situations with a certain collision risk and rated its criticalities. Banet and Bellet (2008) observed that the criticality ratings of car drivers were significantly higher than those of motorcyclists. They assumed that the latter have a more pronounced sensation-seeking personality, hence, they accept more critical driving situations as tolerable. This demonstrates that the collected evaluations in studies 1, 2, and 3 might have been affected by the individual characteristics of the participants.

Sixth, numerical values are commonly used to process the evaluations and to enable a statistical analysis (Moosbrugger & Kelava, 2020). In this case, originally ordinal scaled variables are treated as interval scaled. This is problematic because the assumptions for the application of these analysis methods are not met (Moosbrugger & Kelava, 2020). However, Moosbrugger and Kevala (2020) stated that ratings with a response format of five or more scale levels can be interpreted as approximate interval scaled. This applies to both tested scales.

To conclude, evaluations obtained with scales do not necessarily correspond to the perception of the driving situations even when they are validated. First, responses may be biased for several reasons. Second, the statistical analysis may be based on wrong assumptions. However, some of these aspects may be counteracted by different design methods or neglected due to their presumably weak influence. Other assessment tools may allow a continuous and less biased evaluation of the criticality of driving situations, but they are not validated yet.

3.1.2 Repeatedly taking over

The results of study 1 indicate that effort and behavior barely change when repeatedly taking over vehicle control. Except brake inputs became weaker with increasing experience. This is in contrast to Brandenburg and Roche (2020), who observed an increase of deceleration over three trials. Visually inspecting the brake pedal input (see Figure 8f in appendix C) shows that the mean values per trial vary, indicating no specific trend. Additionally, the marginal coefficient of determination of this model was very low, suggesting that the factor ‘trial’ did not explain much variance. It seems like a linear regression does not approximate the change of braking adequately.

Perceived effort ratings, TOTs, and maximal steering wheel angles remained stable over trials. As mentioned in section 2.1.3, the change of criticality ratings are not considered further because they were assessed with a scale whose reliability has not yet been investigated. There are two possible reasons for the missing effects of the dependent variables. First, fatigue likely set in over trials which is assumed to lead to higher effort to stay awake (de Waard, 2002). At the same time, participants became more trained which probably made it easier for them to take over. Hence, the increased practice may have compensated for increasing effort to stay alert. Second, subjective criticality and take-over behavior might have changed during the five training trials. Or they adopted within the first experimental ones due to increasing practice with the driving simulator and the take-over scenario and hardly changed after that. A phenomenon like the one described was observed by Forster et al. (2019). They showed that take-over performance concerning transitions to SAE-level 2 or 3 fits a learning curve with a steep increase within the first trials followed by a stabilization (Forster et al., 2019). This curve is described by the *power law of practice* which considers training as mandatory to improve performance, i.e. react faster or make fewer errors, resulting in a noticeable improvement within the first repetitions (Neves & Anderson, 1981; Newell & Rosenbloom, 1981). But the change aims toward zero from a certain point of practice (Neves & Anderson, 1981). The first part of this assumption is supported by the results of different driving simulator studies. The participants were requested to take over two, three, or six times, respectively. It was observed that subjective criticality and perceived effort decreased, trust increased, take-overs became faster, and minimal TTCs increased over the repeated take-overs (Gold et al., 2015; Hergeth et al., 2015, 2017; Payre et al., 2016; Roche et al., 2018). Interestingly, deceleration increased in Brandenburg and Roche (2020) while it decreased Hergeth et al. (2017). In study 1, a larger number of trials ($N = 10$) compared to previous studies ($N = 2, 3, 6$) was conducted. Here, the ratings and behavior might have changed in the first trials and stabilized afterwards. For the analysis, a linear trend was adopted which might not have been appropriate to represent this effect.

The findings on repeated take-overs are promising because they suggest that the results of driving simulator studies that examine several take-overs in a row are relatively valid because perceived effort and behavior do not change with increasing practice except maybe in the first take-overs. Using such designs might be necessary for several reasons: test efficiency, limited pool of participants, or testing within-subjects to control for individual differences between the experimental groups. These are valid reasons to take the risk of small behavioral changes when testing several trials per participant. Nevertheless, it remains essential to systematically balance the experimental conditions to avoid any order effect. Researchers should also stay aware of potential behavioral changes when investigating several take-overs in one session.

3.1.3 Effects of time budget and vehicle dynamics

The effects of time budget and vehicle dynamics on criticality ratings, take-over behavior, and performance were the focus of the present thesis. They were used to evaluate how drivers perceive, behave, and perform (in) time-critical and dynamic driving situations to answer the research question *what might happen if drivers take over vehicle control in time-critical and dynamic driving situations*.

Kahneman's capacity model of attention (1973) was used to explain the observed interactions between time budget and vehicle dynamics in the lane change situations (study 3). It assumes that the human cognitive potential is generally limited and must be directed to tasks to be activated. In turn, less capacity is available to carry out further activities. The allocation is determined by different aspects, among them the involuntary attention to novel or sudden stimuli (Kahneman, 1973). Transferred to the investigated lane change situations, it seems that the highly dynamic ones have presented such a trigger because the lateral accelerations were stronger and a loss of control was more likely. Also, the visual impression of the dynamic lane changes was intense. Thus, these aspects were processed, while fewer cognitive capacities were available to process other stimuli such as time budget. Consequently, the time budgets could hardly affect the dependent variables. It can be assumed that system 1 described by Kahneman (2011) was strongly involved in these situations. It is one of the two ways the brain works and is responsible for intuitive, fast thinking. System 1 requires little or no effort and operates automatically and involuntarily. In contrast, system 2 is associated with slow, effortful reasoning and concentration. It draws attention to voluntary, cognitively demanding tasks (Kahneman, 2011). With lower vehicle dynamics, the lateral accelerations were smaller. Hence, the situations might have appeared less threatening. In these cases, more capacities were available to pay attention to other aspects. So, the time budget could affect the dependent variables. System 1 might have been less involved.

Criticality ratings

The present thesis showed that drivers can discriminate between different degrees of criticality of driving situations realized by time budget and vehicle dynamics – assuming that the ratings correspond to the driving situation and do not reflect the outcome of the maneuver.

The effects of time budget are in line with previous studies which demonstrated that lower time budgets lead to higher criticality ratings (Roche & Brandenburg, 2018, 2020).

The results of the present thesis extend the previously available knowledge that participants are sensitive to low time budgets and vehicle dynamics. This is a noteworthy finding as it indicates that drivers do not only perceive temporal differences in driving situations but also very small ones and variations of vehicle dynamics. The insights generated by the criticality ratings support the interpretation of observed take-over behavior.

Take-over behavior

Take-over behavior was evaluated by TOTs, deceleration, and steering variables. The observed means and standard deviations of TOTs for each study are presented in Table 1. Surprisingly, the means were lowest in the first study, even though, these situations were rated as least critical. Rather, it was expected that participants would take over slower if they evaluated the situations as less critical. This disparity may be due to the different simulators and software that was used in study 1 compared to studies 2 and 3. The setting and sampling frequency of study 1 may have led to the faster take-overs. But mean TOTs in study 2 were also lower than in study 3, even though, they were rated as less critical. Studies 2 and 3 were conducted in the same driving simulator but with different participants. These results indicate that comparing the absolute criticality ratings between studies and across different respondents might be misleading. Comparing criticality ratings within one study and with the same set of participants seems to be adequate as indicated by the higher TOTs and lower criticality ratings in the second lane changes compared to the first.

Previous studies are consulted to classify the magnitude of the observed TOT-values. Eriksson and Stanton (2017) reviewed 25 driving simulator studies concerning time budgets and TOTs. They identified a mean TOT of 2,690 ms (*SD* = 1,960 ms) ranging from 1,550 s to 15,000 ms (Eriksson & Stanton, 2017). This shows that the take-overs observed in the present thesis are very fast. There are three possible reasons for this finding.

First, the lower time budgets in the present thesis compared to the studies included in the review of Eriksson and Stanton (2017) could be the reason for the difference. These lower time budgets may have elicited the faster take-overs as observed in Zhang et al. (2019).

Second, the participants in studies 1, 2, and 3 did not perform any non-driving related task (NDRT) and could focus on the environment. Whereas, in most of the studies reported by Eriksson and Stanton (2017), participants were distracted by a NDRT. The missing distraction might have led to faster switches from supervising to controlling the vehicle and, hence, to lower TOTs. Indeed, Feldhütter et al. (2017) observed faster take-overs of non-distracted drivers compared to those performing a NDRT.

Third, in the present thesis, participants were highly trained to take over vehicle control before the experimental session started. Whereas participants of studies included in the review of Eriksson and Stanton (2017) experienced no or very little training and fewer experimental trials, e.g. Gold et al. (2016) or Lorenz et al. (2014). As discussed in the section on repeated take-overs (see 3.1.2), it could be that TOTs decreased within the first training and experimental trials due to increasing practice and were, then, on a lower level.

Table 1: Means and standard deviations (SD) of criticality ratings and take-over times (TOTs) per study.

Study	**Mean criticality ratings [pts.]**	**SD criticality ratings [pts.]**	**Mean TOTs [ms]**	**SD TOTs [ms]**
Study 1: Lane change situations	41.44	21.99	543	283
Study 2: Brake situations	49.64	26.16	567	333
Study 3: First lane change situations	65.42	25.81	825	262
Study 3: Second lane change situations	56.45	26.05	907	305

Study 3 demonstrated that TOTs were only affected by the time budget in the second lane change. Participants took over more slowly when they had the time to do so. With lower time budgets, they reacted appropriately by taking over faster. Roche and Brandenburg (2018, 2020) observed comparable effects. Similarly, Zhang et al. (2019) extracted in their meta-analysis that higher time budgets led to slower take-overs. This indicates that drivers may recognize when they have more time and use it to initiate the take-over. Whereas they react as fast as possible when necessary. A reason that the effect was not found in the first lane change might be its lower predictability. The lead vehicle covered the first obstacle while the second one was earlier visible. Hence, the first take-over situation appeared more sudden and likely produced faster reactions irrespective of time budget. This assumption is supported by the descriptively lower mean TOTs in the first compared to the second lane change (see Table 1). A floor effect could be the consequence. Therefore, different time budgets would have little or no influence on TOTs in the first situation.

More critical and dynamic situations were evaluated adequately, as indicated by the criticality ratings. It seems as if this perception elicited stronger decelerations and steering. Roche and Brandenburg (2018, 2020) also observed that participants decelerated and steered more extreme when having less time to take over. These effects may have the following two reasons. First, with more critical time budgets, participants had less time to process the driving environment and select an appropriate action. Due to that, participants were probably frightened and reacted with stronger deceleration and steering as suggested by Davis (1984). However, such behavior was not necessary to avoid the obstacles. Second, in the more dynamic brake situations, the merging vehicle decelerated stronger. Hence, extreme deceleration was necessary and adequate to avoid a collision. Similarly, larger steering in the more dynamic lane change situations was required because the trajectories to evade the obstacles were steeper.

Only in the second lane change, time budget did not affect maximal decelerations. These were the situations in which time budget affected TOTs. Hence, it could be that the faster take-overs reduced the necessity to brake stronger. While in the brake situations and first lane changes, participants compensated for the missing differences of TOTs by more extreme braking.

It can be summarized that more critical time budgets and vehicle dynamics act as intense and adverse stimuli and elicit the corresponding responses (Davis, 1984). Similarly, the predictability of the take-overs is suspected to affect the intensity of take-over actions. Additionally, high vehicle dynamics seem to strongly draw attention, activating system 1, leaving little cognitive capacities to process other stimuli such as time budget (Kahneman, 1973).

Take-over performance

Deceleration and steering should be strong enough to avoid collisions with the merging vehicle or the obstacles. But they should be as weak as possible to avoid loss of vehicle control, lane departures, and rear-end collisions. That way, the respective automation in studies 2 and 3 was designed. Hence, the deceleration and steering in comparison to the automation and the potential consequences, e.g. lane changes, lane departures, and collisions, were used to evaluate take-over performance.

The comparison with the automation is a rather novel approach to evaluate performance. The findings from this are discussed here. In the beginning of the research project, it was unclear which dependent variable fits best. In study 2, the difference between the deceleration behavior of the participants and the automation was calculated. The effects of time budget and vehicle dynamics on the deviation were then analyzed with linear mixed-effects models. In study 3, t-tests between the participants' deceleration and steering and the corresponding values of the automation were calculated for each condition. To evaluate take-over performance, t-tests were identified to be better suited because they reveal potential overreactions per experimental condition. This was more useful than investigating the effects of time budget and vehicle dynamics. However, for the evaluation of the performance, the behavior of the ADS served as reference. This is questionable because the ADS could have been implemented in many ways resulting in different data. For future comparisons, it would be advisable to create a truly ideal reference that meets criteria yet to be defined.

The results of study 2 suggest that participants decelerated strongly almost irrespective of the vehicle dynamics. Similarly, in study 3, participants braked under all conditions. This behavior can be interpreted as an overreaction because no deceleration was necessary in study 3. Participants might have considered strong braking as inevitable to avoid a collision with the merging vehicles or the obstacles. Or they might have preferred to reduce speed for the lane change maneuver. Either way, strong deceleration is hazardous because it may result in rear-end collisions and loss of vehicle control.

In the more time-critical and dynamic brake situations (study 2), more lane changes occurred. Possibly, changing lanes was the participants' strategy to avoid a collision when less time for braking was available or when the situation was more dynamic. This strategy was successful to some extent: Only three collisions (out of 357 trials) occurred in study 2. However, it could be that while avoiding the threat to collide with the merging vehicle, participants did not carefully check the lane they were changing to. In turn, participants risked collisions with vehicles on the adjacent lanes. Unfortunately, these events were neither simulated nor investigated in study 2. Recording and analyzing the gaze behavior would have provided more insights here.

In study 3, more lane departures occurred when drivers had less time to avoid a collision. In these situations, participants steered more than necessary. Possibly, the stronger steering led to lane departures, demonstrating that a take-over under these conditions might be dangerous. Surprisingly, the lane departure frequencies were not affected by vehicle dynamics. It seems that participants were able to maintain control despite the stronger steering with higher vehicle dynamics.

In the first lane changes of study 3, 30 collisions took place. Surprisingly, in 26 trials, participants collided with the first obstacle when they steered back to the left lane too early. This can be due to the participants' striving to return to the initial state (here the lane) and to avoid the second obstacle as fast as possible without carefully checking their surroundings. It could also be that participants failed to develop an appropriate representation of the vehicle and the environment because the study was conducted in a driving simulator. The extension of the body schema to the nearly static vehicle may have been incomplete. Already in 1911, Head and Holmes identified passive movement as a powerful method to generate a correct body schema (Head & Holmes, 1911). The movement of the motion seat might not have been sufficient for that. Besides, it can be assumed that the vehicle in the driving simulator was not fully accepted as a *tool,* respectively extension of the body. This phenomenon was reported by Maravita and Iriki (2004), although in a very different setting. The researchers observed that the neuronal activity of primates changed after using a rake to get food. The interpretation of their finding was that the body schema extended through tool use (Maravita & Iriki, 2004). Based on that, it can be inferred that the experience with the simulated vehicle was insufficient to evoke an extension of the body schema. In real traffic, it should be easier for the drivers because the vehicle is truly moving. Consequently, it can be assumed that they would show different behavior with fewer collisions when steering back to the initial lane.

The evaluation of take-over performance demonstrated that participants decelerate too much and steer too strong when taking over in such highly time-critical and dynamic driving situations. This behavior is hazardous because too high decelerations and strong steering increase the risk of losing vehicle control when the maximal horizontal force is exceeded.

Besides, rear-end collisions, collisions with vehicles on the adjacent lanes, and lane departures are more likely. The observed consequences of take-overs, e.g. collisions, are noteworthy because none of them would have happened if the drivers had not taken over.

3.2 Development of assistant systems for take-overs

The presented research has shown that driver-initiated take-overs can have serious consequences. However, drivers always have the right to override the system (United Nations Economic Commission for Europe, 2014). Hence, ways need to be found to enable such take-overs while their harmful consequences are prevented. One way is to impede or moderate the drivers' actions by assistance systems. It is assumed that the application of such systems improves the take-over quality and results in safer take-overs (Nguyen & Müller, 2020).

Researchers introduced different concepts for assistance systems. Wada et al. (2016) and Saito et al. (2018) suggested a control transition with a smooth transfer, the so-called *shared authority* mode. During the take-over, the vehicle control should be gradually handed from the ADS to the driver while the system may adjust drivers' extreme steering. This would grant the drivers more time to gain situation awareness. Walch et al. (2015) proposed different types of procedures for transferring the control from the ADS to the driver. One type is the *immediate hand-over* with a complete control shift once the system detects a driver input. This is the currently widely implemented form of hand-over procedure. Another type suggested by Walch et al. (2015) ought to assist the driver during the take-over, the *system-monitored hand-over*. The idea of this type is that the system adjusts the drivers' input in case it would result in a critical situation (Walch et al., 2015). This procedure may support the driver during the take-over and as long as needed after it.

In the course of the DFG-research project, the automotive engineers from the Technische Universität Berlin developed two different assistance systems that may operate in brake and lane change situations (Nguyen & Müller, 2020). The first is based on a *threshold* and is a new concept. Drivers have to overcome a certain value before the ADS is deactivated and their input is executed (Nguyen & Müller, 2020). This threshold is lower for low critical and higher for highly critical driving situations. For example, more brake pedal force is needed to deactivate the ADS when time budgets are lower or TU is higher. With the second assistance system, the *cooperative* assistant, driver input is possible at all times, but the needed momentum to realize steering or decelerating is higher when the situation is more critical (Nguyen & Müller, 2020). Hence, steering demands more torque when the situation is more dynamic, i.e. when such behavior is hazardous because a loss of vehicle control is likely. This concept is similar to the suggested one by Walch et al. (2015).

Nguyen and Müller (2020) investigated which of their implementations, the *threshold* and *cooperative* assistant, performed best in brake and lane change situations. In a driving simulator, five participants experienced the most critical brake and lane change situations from studies 2 and 3, three times. In each trial, one of the two versions of assistance or no assistance was activated. Nguyen and Müller (2020) visualized the timely course of time budgets for the brake situations and the lateral position for the lane change situations. They compared it to the course of the automation. The visual inspection of the timely courses indicated that participants with the cooperative assistant deviated less from the automation's time budgets or path for both situations than participants with the threshold or without any assistance (Nguyen & Müller, 2020).

Their results give a first impression that the cooperative assistant might be superior. However, special configurations of the two assistance systems were implemented with specific thresholds and settings based on time budget and TU (Nguyen & Müller, 2020). It is questionable whether these implementations of the assistance systems would prevent lane departures or collisions, especially those caused by too early steering back to the initial lane as observed in study 3. Different realizations of the two concepts might have shown different results. Furthermore, the versions were tested with only a small number of participants in one experimental condition per scenario. Besides, the results were not statistically analyzed but visually inspected. More elaborate testing and investigating different implementations of the assistance systems are required to consolidate the findings. Moreover, examining the implementations with more participants in a more realistic environment, e.g. a test track, would enhance the generalizability of the results. Such a study was conducted on a former military airfield in February and March 2020 in the course of the DFG-research project (Brandenburg et al., in preparation).

3.3 Limitations and future studies

The following limitations concerning all studies should be considered. They contribute to ideas for future investigations.

First, the rating scales were validated for take-over situations in lane change maneuvers only. Besides, the criticality was varied by time budgets only. Further scenarios and different variations of the criticality of the driving situations might affect the ratings differently. Hence, it would enhance the generality of the results if the scales were validated in additional scenarios.

Second, participants were triggered to take over in studies 2 and 3. This procedure was necessary to elicit take-overs when the intended values of time budget and TU were met. It enabled the assignment of trials to experimental conditions and the comparability between them. However, it should be noted that the intended values were met at the moment of the trigger. They differed when participants took over because it could not be controlled when they did so. Besides, the take-over situations were artificially created. Hence, I cannot claim that I investigated naturalistic driver-initiated take-overs. However, the results indicate *what might* happen if drivers took over in these driving situations. This should be noted when transferring the results to driver-initiated take-overs. Future studies should focus on take-overs truly initiated by the drivers to be able to observe realistic take-over behavior.

Third, the studies were conducted in fixed-based driving simulators. In study 1, the simulator consisted of a vehicle mock-up. For studies 2 and 3, a high-fidelity simulator was used with a motion seat to induce lateral and longitudinal accelerations. It is questionable how realistic these movements were. This might limit the generalizability of the results. For example, it could be that the many observed collisions in study 3 were a by-product of the driving simulator. As discussed in section 3.1.3 on take-over performance, the participants might have failed to generate an appropriate representation of the vehicle. However, investigating such highly critical situations on a test track and putting the drivers at high risks would have been irresponsible. Besides, Eriksson et al. (2017) showed that driving simulators are highly valid to investigate workload, perceived usefulness, satisfaction, and behavior in automated driving. And yet, in the course of this DFG-research project, a study was conducted on a test track (Brandenburg et al., in preparation). Participants were operating a real, automated vehicle. Virtual obstacles and vehicles were displayed on the drivers' track. Similar to studies 2 and 3, the in-built automation would have been able to evade the displayed obstacles. But participants were again triggered to take over. The results are not available yet, but they will show whether participants behave similarly in more realistic environments and with more pronounced accelerations.

Fourth, several take-overs were tested in a row. This was due to test efficiency and the experimental setting. As presented in the introduction, increased practice in taking over may influence experience and behavior. But the results of study 1 indicated that perceived effort and behavior barely change over trials. Hence, it can be concluded that the results of studies 2 and 3 are valid, even though, several take-overs were tested in one session. With the findings from study 1, it would have even been justifiable to exclude the first trials from the analysis of studies 2 and 3 because the behavior stabilizes after the first take-overs (see section 3.1.2).

The effects of individual parameters were not investigated in the present thesis. Previous studies demonstrated that characteristics such as drivers' personality (Banet & Bellet, 2008; Mesken et al., 2007; Zeeb et al., 2015), age (Körber et al., 2016), or multitasking ability (Körber et al., 2015) influence ratings and take-over behavior. Hence, open questions remain, such as: How do drivers' characteristics influence criticality ratings and take-over behavior in highly time-critical and dynamic driving situations?

Another aspect in this context has not yet been dealt with in this thesis. These are the relations between subjective criticality and take-over behavior and performance. On the one hand, a certain perception of the driving situation likely leads to corresponding behavior and performance, e.g. a driver who evaluates the driving situation as very critical might brake stronger than a driver who perceives it as less critical. Hence, an effect of subjective criticality on take-over behavior and performance is possible. The additional analysis in study 3 partially supported this assumption by showing that behavior in terms of TOTs correlated with criticality ratings. On the other hand, as discussed earlier, the timing of the evaluation is crucial. Behavior resulting in a certain performance might lead to corresponding criticality ratings if they were collected after the experience, e.g. situations that resulted in collisions might be rated as more critical. Indeed, study 3 showed that the criticality ratings also correlated with collision frequency. Hence, the explanatory value of retrospective evaluations concerning take-over behavior is limited. Investigating this relation may indicate whether the timing of the assessment of criticality is crucial.

3.4 Conclusion and outlook

The present thesis focused on the effects of driver-initiated take-overs on criticality ratings, take-over behavior, and performance. Previously, this type of control transfer had received very little attention. The research question of the thesis can be answered as follows: Drivers perceive more time-critical and dynamic driving situations as more critical and show extreme take-over behavior in terms of TOTs, steering, and deceleration. These actions result in severe consequences: loss of vehicle control, unintentional lane changes, lane departures, and collisions with the obstacles, vehicles traveling on the adjacent lane, or the crash barrier. Hence, it can be concluded that driver-initiated take-overs can carry a high risk.

A final search shortly before submitting this thesis revealed that no new studies on the topic mentioned had been published, other than those that arose as part of this research project. And this despite the fact that the presented studies showed that there is a high need to deal more intensively with driver-initiated take-overs. Hence, research should focus on this aspect of automated driving. Besides, ways must be found and investigated to prevent the serious consequences, especially as long as the legal regulations require the possibility of driver intervention at any times. And there are currently no indications that the Vienna convention will be amended concerning the provisions on automated driving soon.

Only recently, in March 2022, the first Tesla factory in Europe was opened in Brandenburg, Germany (Hahn, 2022). Tesla is mainly known for its pioneering role in electric mobility. But the built-in systems also enable automated driving at SAE level 2. In addition, the company is already equipping its vehicles with hardware that can one day enable fully automated driving (SAE-level 5) through software updates (Tesla, Inc., 2022). This shows that the automotive industry is preparing for higher levels of automation, while the technical and regulatory requirements are not yet in place. This gap should be closed as soon as possible to be able to benefit from the advantages of automated driving introduced in section 1.

4 References

AAA Foundation for Traffic Safety. (2021). *New American Driving Survey: Updated Methodology and results from July 2019 to June 2020*.

Agrawal, S., & Peeta, S. (2021). Evaluating the impacts of situational awareness and mental stress on takeover performance under conditional automation. *Transportation Research Part F: Traffic Psychology and Behaviour*, *83*, 210–225. https://doi.org/10.1016/j.trf.2021.10.002

Anderson, J. M., Kalra, N., Stanley, K. D., Sorensen, P., Samaras, C., & Oluwatola, O. A. (2014). *Autonomous vehicle technology: A guide for policymakers*. Rand Corporation.

Banet, A., & Bellet, T. (2008). Risk awareness and criticality assessment of driving situations: A comparative study between motorcyclists and car drivers. *IET Intelligent Transport Systems*, *2*(4), 241. https://doi.org/10.1049/iet-its:20080037

Becker, S., Brandenburg, S., & Thüring, M. (2022). Driver-initiated take-overs during critical braking maneuvers in automated driving—The role of time headway, traction usage, and trust in automation. *Accident Analysis and Prevention*.

BMW. (2020). *Autonomous Driving*. /content/grpw/websites/bmwgroup_com/en/innovation/technologies-and-mobility/autonomes-fahren.html

Boggs, A. M., Arvin, R., & Khattak, A. J. (2020). Exploring the who, what, when, where, and why of automated vehicle disengagements. *Accident Analysis & Prevention*, *136*, 105406. https://doi.org/10.1016/j.aap.2019.105406

Bogner, K., & Landrock, U. (2016). Response Biases in Standardised Surveys. *GESIS Survey Guidelines*. https://doi.org/10.15465/GESIS-SG_EN_016

Borojeni, S. S., Chuang, L., Heuten, W., & Boll, S. (2016). Assisting Drivers with Ambient Take-Over Requests in Highly Automated Driving. *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular*

Applications, 237–244. https://doi.org/10.1145/3003715.3005409

Brandenburg, S., Becker, S., Trukenbrod, A. K., & Thüring, M. (in preparation). *A field study on take-overs during highly-dynamic lane changes in automated driving*.

Brandenburg, S., & Roche, F. (2020). Behavioral Changes to Repeated Takeovers in Automated Driving: The Drivers' Ability to Transfer Knowledge and the Effects of Takeover Request Process. *Transportation Research Part F: Psychology and Behaviour*. https://doi.org/10.1016/j.trf.2020.06.002

Breuer, S., & Rohrbach-Kerl, A. (2015). *Fahrzeugdynamik*. Springer Vieweg. https://doi.org/10.1007/978-3-658-09475-1_1

Casner, S. M., Hutchins, E. L., & Norman, D. (2016). The challenges of partially automated driving. *Communications of the ACM*, *59*(5), 70–77. https://doi.org/10.1145/2830565

Damböck, D., Farid, M., Tönert, L., & Bengler, K. (2012). Übernahmezeiten beim hochautomatisierten Fahren. *Tagung Fahrerassistenz*, 16–28.

Davis, M. (1984). The Mammalian Startle Response. In *Neural Mechanisms of Startle Behavior* (pp. 287–351). Springer. https://link.springer.com/chapter/10.1007/978-1-4899-2286-1_10

de Waard, D. (2002). Mental Workload. In *Human Factors for Highway Engineers* (1st ed., pp. 161–175). Bingley. https://trid.trb.org/view/708725

Department for Transport. (2015). *The pathway to Driverless Cars* (p. 40). Department for Transport.

Destatis. (2021). *Verkehrsunfälle*. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/_inhalt.html

Du, N., Zhou, F., Pulver, E. M., Tilbury, D. M., Robert, L. P., Pradhan, A. K., & Yang, X. J. (2020). Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving. *Transportation Research Part C:*

Emerging Technologies, *112*, 78–87. https://doi.org/10.1016/j.trc.2020.01.006

Eick, E.-M., & Debus, G. (2005). Adaptation effects in an automated car following scenario. *Traffic and Transport Psychology: Theory and Application*, 243–255.

Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden* (5.). Beltz.

Epple, S., & Brandenburg, S. (2018). *Driver-Initiated Take-Over Behavior in Critical Driving Situations in Highly Automated Driving*. Human Factors and Ergonomics Society Europe Chapter Annual Meeting, Berlin.

Eriksson, A., Banks, V. A., & Stanton, N. A. (2017). Transition to manual: Comparing simulator with on-road control transitions. *Accident Analysis & Prevention*, *102*, 227–234. https://doi.org/10.1016/j.aap.2017.03.011

Eriksson, A., & Stanton, N. A. (2017). Takeover Time in Highly Automated Vehicles: Noncritical Transitions to and from Manual Control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*(4), 689–705. https://doi.org/10.1177/0018720816685832

ERTRAC. (2015). *Automated Driving Roadmap Technical Report of the European Road Transport Research Advisory Council*. European Union. http://www.ertrac.org/uploads/documentsearch/id38/ERTRAC_Automated-Driving-2015.pdf

European Commission. (2016). *Intelligent Transport Sytems: Road* [Text]. Mobility and Transport - European Commission. https://ec.europa.eu/transport/themes/its/road_it

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, *77*, 167–181. https://doi.org/10.1016/j.tra.2015.04.003

Feldhütter, A., Gold, C., Schneider, S., & Bengler, K. (2017). How the Duration of Automated Driving Influences Take-Over Performance and Gaze Behavior. In C. M. Schlick, S. Duckwitz, F. Flemisch, M. Frenz, S. Kuz, A. Mertens, & S. Mütze-

Niewöhner (Eds.), *Advances in Ergonomic Design of Systems, Products and Processes* (pp. 309–318). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-53305-5_22

Feldhütter, A., Kroll, D., & Bengler, K. (2018). Wake Up and Take Over! The Effect of Fatigue on the Take-over Performance in Conditionally Automated Driving. *Proceedings of 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2080–2085. https://doi.org/10.1109/ITSC.2018.8569545

Forster, Y., Hergeth, S., Naujoks, F., Beggiato, M., Krems, J. F., & Keinath, A. (2019). Learning to use automation: Behavioral changes in interaction with automated driving systems. *Transportation Research Part F: Traffic Psychology and Behaviour*, *62*, 599–614. https://doi.org/10.1016/j.trf.2019.02.013

Forster, Y., Naujoks, F., Neukum, A., & Huestegge, L. (2017). Driver compliance to take-over requests with different auditory outputs in conditional automation. *Accident Analysis & Prevention*, *109*, 18–28. https://doi.org/10.1016/j.aap.2017.09.019

Frey, B. B. (2018). Standards for Educational and Psychological Testing. In *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE Publications, Inc. https://doi.org/10.4135/9781506326139.n662

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, *7*(3), 385–400. https://doi.org/10.1016/0191-8869(86)90014-0

Gasser, T. M., Arzt, C., Ayoubi, M., Bartels, A., Bürkle, L., Eier, J., Flemisch, F., Häcker, D., Hesse, T., Huber, W., Lotz, C., Maurer, M., Ruth-Schumacher, S., Schwarz, J., & Vogt, W. (2012). *Rechtsfolgen zunehmender Fahrzeugautomatisierung: Gemeinsamer Schlussbericht der Projektgruppe* (Bundesanstalt für Straßenwesen, Ed.). Wirtschaftsverlag NW.

General Motors. (2020). *2018 Self-Driving Safety Report*. https://www.gm.com/content/dam/company/docs/us/en/gmcom/gmsafetyreport.pdf

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. https://doi.org/10.1177/1541931213571433

Gold, C., Körber, M., Hohenberger, C., Lechner, D., & Bengler, K. (2015). Trust in Automation – Before and After the Experience of Take-over Scenarios in a Highly Automated Vehicle. *Procedia Manufacturing*, *3*, 3025–3032. https://doi.org/10.1016/j.promfg.2015.07.847

Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking Over Control From Highly Automated Vehicles in Complex Traffic Situations: The Role of Traffic Density. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(4), 642–652. https://doi.org/10.1177/0018720816634226

Hahn, A. (2022, March 22). *Tesla-Werk in Brandenburg eröffnet*. https://www.tagesschau.de/wirtschaft/unternehmen/tesla-fertigung-brandenburg-101.html

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Th ANNUAL MEETING*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, *52*, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hartig, J., Frey, A., & Jude, N. (2008). Validität. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion*. Springer.

Head, H., & Holmes, G. (1911). SENSORY DISTURBANCES FROM CEREBRAL LESIONS. *Brain*, *34*(2–3), 102–254. https://doi.org/10.1093/brain/34.2-3.102

Heller, O. (1982). *Theorie und Praxis des Verfahrens der Kategorieunterteilung (KU)* (pp. 1–15). Psychologisches Institut der Universität Würzburg.

Hergeth, S., Lorenz, L., & Krems, J. F. (2017). Prior Familiarization With Takeover Requests Affects Drivers' Takeover Performance and Automation Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*(3), 457–470. https://doi.org/10.1177/0018720816678714

Hergeth, S., Lorenz, L., Krems, J. F., & Toenert, L. (2015). Effects of Take-Over Requests and Cultural Background on Automation Trust in Highly Automated Driving. *Proceedings of 8th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, 331–337. http://drivingassessment.uiowa.edu/sites/default/files/DA2015/papers/051.pdf

Herrmann, S., Utschick, W., Botsch, M., & Keck, F. (2015). Supervised Learning via Optimal Control Labeling for Criticality Classification in Vehicle Active Safety. *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 2024–2031. https://doi.org/10.1109/ITSC.2015.328

Hu, Q., Guo, G., Wang, H., Tang, M., Lv, C., & Cao, D. (2019). How to Define the Passenger's Hazard Perception Level by Combining Subjective and Objective Measures? In C. Stephanidis (Ed.), *HCI International 2019 – Late Breaking Papers* (Vol. 11786, pp. 40–53). Springer International Publishing. https://doi.org/10.1007/978-3-030-30033-3_4

Hyundai. (2020). *An Autonomous Driving Future*. https://www.hyundai.com/au/en/why-hyundai/autonomous-driving

ISO 31000:2009. (2009). *Risk management—Guidelines*. https://www.iso.org/obp/ui#iso:std:iso:31000:ed-2:v1:en

Junietz, P., Schneider, J., & Winner, H. (2017). *Metrik zur Bewertung der Kritikalität von Verkehrssituationen und -szenarien*. 12.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.398.5285&rep=rep1&type=pdf

Kahneman, D. (2011). *Thinking fast and slow*. Farrer, Straus & Giroux.

Kamm, W. (1936). *Das Kraftfahrzeug: Betriebsgrundlagen, Berechnung, Gestaltung und Versuch*. Springer.

Kerschbaum, P., Lorenz, L., & Bengler, K. (2015). A transforming steering wheel for highly automated cars. *2015 IEEE Intelligent Vehicles Symposium (IV)*, 1287–1292. https://doi.org/10.1109/IVS.2015.7225893

Kesting, A., Treiber, M., Schönhof, M., & Helbing, D. (2008). Adaptive cruise control design for active congestion avoidance. *Transportation Research Part C: Emerging Technologies*, *16*(6), 668–683. https://doi.org/10.1016/j.trc.2007.12.004

Körber, M., Gold, C., Lechner, D., & Bengler, K. (2016). The influence of age on the take-over of vehicle control in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *39*, 19–32. https://doi.org/10.1016/j.trf.2016.03.002

Körber, M., Weißgerber, T., Blaschke, C., Farid, M., & Kalb, L. (2015). Prediction of take-over time in highly automated driving by two psychometric tests. *DYNA*, *82*(193), 195–201. https://doi.org/10.15446/dyna.v82n193.53496

Lorenz, L., Kerschbaum, P., & Schumann, J. (2014). Designing take over scenarios for automated driving: How does augmented reality support the driver to get back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 1681–1685. https://doi.org/10.1177/1541931214581351

Lu, Z., Happee, R., Cabrall, C., Kyriakidis, M., & de Winter, J. C. F. (2016). Human factors of transitions in automated driving: A general framework and literature survey. *Transportation Research Part F: Traffic Psychology and Behaviour*, *43*, 183–198. https://doi.org/10.1016/j.trf.2016.10.007

Lv, C., Cao, D., Zhao, Y., Auger, D. J., Sullman, M., Wang, H., Dutka, L. M., Skrypchuk, L., & Mouzakitis, A. (2018). Analysis of autopilot disengagements occurring during autonomous vehicle testing. *IEEE/CAA Journal of Automatica Sinica*, *5*(1), 58–68. https://doi.org/10.1109/JAS.2017.7510745

Ma, S., Zhang, W., Yang, Z., Kang, C., Wu, C., Chai, C., Shi, J., Zeng, Y., & Li, H. (2021). Take over Gradually in Conditional Automated Driving: The Effect of Two-stage Warning Systems on Situation Awareness, Driving Stress, Takeover Performance, and Acceptance. *International Journal of Human–Computer Interaction*, *37*(4), 352–362. https://doi.org/10.1080/10447318.2020.1860514

Maggi, D., Romano, R., & Carsten, O. (2020). Transitions Between Highly Automated and Longitudinally Assisted Driving: The Role of the Initiator in the Fight for Authority. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 001872082094618. https://doi.org/10.1177/0018720820946183

Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences*, *8*(2), 79–86. https://doi.org/10.1016/j.tics.2003.12.008

Martens, Pauwelussen, Schieben, Flemisch, Merat, Jamson, & Caci. (2007). *Human Factors' aspects in automated and semi-automatic transport systems: State of the art* (3.2.1; p. 55). European Commission.

McCall, R., McGee, F., Meschtscherjakov, A., Louveton, N., & Engel, T. (2016). Towards A Taxonomy of Autonomous Vehicle Handover Situations. *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - Automotive'UI 16*, 193–200. https://doi.org/10.1145/3003715.3005456

Melcher, V., Rauh, S., Diederichs, F., Widlroither, H., & Bauer, W. (2015). Take-Over Requests for Automated Driving. *Procedia Manufacturing*, *3*, 2867–2873. https://doi.org/10.1016/j.promfg.2015.07.788

Merat, N., Jamson, A. H., Lai, F. C. H., & Carsten, O. (2012). Highly Automated Driving, Secondary Task Performance, and Driver State. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(5), 762–771. https://doi.org/10.1177/0018720812442087

Mesken, J., Hagenzieker, M. P., Rothengatter, T., & de Waard, D. (2007). Frequency, determinants, and consequences of different drivers' emotions: An on-the-road study using self-reports, (observed) behaviour, and physiology. *Transportation Research Part F: Traffic Psychology and Behaviour*, *10*(6), 458–475. https://doi.org/10.1016/j.trf.2007.05.001

Mok, B. K.-J., Johns, M., Lee, K. J., Ive, H. P., Miller, D., & Ju, W. (2015). Timing of unstructured transitions of control in automated driving. *Intelligent Vehicles Symposium (IV)*, 1167–1172. https://doi.org/10.1109/IVS.2015.7225841

Mok, B. K.-J., Johns, M., Lee, K. J., Miller, D., Sirkin, D., Ive, P., & Ju, W. (2015). Emergency, Automation Off: Unstructured Transition Timing for Distracted Drivers of Automated Vehicles. *Proceedings of the IEEE 18th Intelligent Transportation Systems (ITSC)*, 2458–2464. https://doi.org/10.1109/ITSC.2015.396

Moosbrugger, & Kelava, A. (Eds.). (2020). *Testtheorie und Fragebogenkonstruktion*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-61532-4

Naujoks, F., Mai, C., & Neukum, A. (2014). The effect of urgency of take-over requests during highly automated driving under distraction conditions. *Advances in Human Aspects of Transportation: Part I*. International Conference on Applied Human Factors and Ergonomics, Krakow.

Naujoks, F., & Neukum, A. (2014). Timing of in-vehicle advisory warnings based on cooperative perception. *Proceedings of the Human Factors and Ergonomics Society Europe Chapter Annual Meeting*, 14.

Naujoks, F., Purucker, C., Wiedemann, K., Neukum, A., Wolter, S., & Steiger, R. (2017). Driving performance at lateral system limits during partially automated driving. *Accident Analysis & Prevention*, *108*, 147–162. https://doi.org/10.1016/j.aap.2017.08.027

Neukum, A., & Krüger, H.-P. (2003). Fahrerreaktionen bei Lenksystemstörungen – Untersuchungsmethodik und Bewertungskriterien. *Reifen - Fahrwerk - Fahrbahn.*, *1791*, 297–318.

Neukum, A., Lübbeke, T., Krüger, H. P., Mayser, C., & Steinle, J. (2008). ACC-Stop&Go: Fahrerverhalten an funktionalen Systemgrenzen. *5. Workshop Fahrerassistenzsysteme-FAS*, 141–150.

Neves, D. M., & Anderson, J. R. (1981). Knowledge Compilation: Mechanisms for the Automatization of Cognitive Skills. In *Cognitive Skills and Their Acquisition* (0 ed., pp. 57–84). Psychology Press. https://doi.org/10.4324/9780203728178-7

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In *Cognitive Skills and Their Acquisition* (1st ed., pp. 1–55). Lawrence Erlbaum Associates.

Nguyen, T., & Müller, S. (2020). *Assistance Systems for Driver Interventions in Critical Situations during Automated Driving*. 31th IEEE Intelligent Vehicles Symposium, Las Vegas, USA.

Noy, I. Y., Shinar, D., & Horrey, W. J. (2018). Automated driving: Safety blind spots. *Safety Science*, *102*, 68–78. https://doi.org/10.1016/j.ssci.2017.07.018

Pacejka, H. B. (2006). *Tire Characteristics and Vehicle Handling and Stability* (2nd ed.). Elsevier.

Payre, W., Cestac, J., & Delhomme, P. (2016). Fully Automated Driving: Impact of Trust and Practice on Manual Control Recovery. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(2), 229–241. https://doi.org/10.1177/0018720815612319

Petermeijer, S., Bazilinskyy, P., Bengler, K., & de Winter, J. (2017). Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics*, *62*, 204–215. https://doi.org/10.1016/j.apergo.2017.02.023

Politis, I., Brewster, S., & Pollick, F. (2015). Language-based multimodal displays for the handover of control in autonomous cars. *Proceedings of the 7th International*

Conference on Automotive User Interfaces and Interactive Vehicular Applications, 3–10. https://doi.org/10.1145/2799250.2799262

Politis, I., Revell, K., Stanton, N., Langdon, P., Adebayo, D., Bradley, M., Clarkson, P. J., Skrypchuk, L., Mouzakitis, A., Eriksson, A., & Brown, J. W. H. (2018). An Evaluation of Inclusive Dialogue-Based Interfaces for the Takeover of Control in Autonomous Cars. *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - IUI '18*, 601–606. https://doi.org/10.1145/3172944.3172990

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. *Advances in Human Aspects of Transportation: Part I*, 421–430.

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. *Proceedings of the 58th Human Factors and Ergonomics Society Annual Meeting*, *58*, 2063–2067. https://doi.org/10.1177/1541931214581434

Radlmayr, J., Ratter, M., Feldhütter, A., Körber, M., Prasch, L., Schmidtler, J., Yang, Y., & Bengler, K. (2018). Take-Overs in Level 3 Automated Driving – Proposal of the Take-Over Performance Score (TOPS). In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)* (Vol. 823, pp. 436–446). Springer International Publishing. https://doi.org/10.1007/978-3-319-96074-6_46

Rajamani, R. (2011). *Vehicle Dynamics and Control*. Springer Science & Business Media.

Roche, F. (2021). Assessing subjective criticality of take-over situations after automated driving: Validation of two rating scales. *Accident Analysis & Prevention*, *159*, 1–13. https://doi.org/10.1016/j.aap.2021.106216

Roche, F., Becker, S., & Thüring, M. (2022). What Happens when Drivers of Automated Vehicles Take Over Control in Critical Lane Change Situations? *Transportation*

Research Part F: Psychology and Behaviour, 407–422. https://doi.org/10.1016/j.trf.2021.11.021

Roche, F., & Brandenburg, S. (2018). Should the urgency of an auditory-tactile takeover request match the situational criticality? *Proceedings 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 1035–1040. https://doi.org/10.1109/ITSC.2018.8569650

Roche, F., & Brandenburg, S. (2020). Should the urgency of visual-tactile takeover requests match the criticality of takeover situations? *IEEE Transactions on Intelligent Vehicles*, *5*(2), 306–313. https://doi.org/10.1109/TIV.2019.2955906

Roche, F., Somieski, A., & Brandenburg, S. (2018). Behavioral Changes to Repeated Takeovers in Highly Automated Driving: Effects of the Takeover Request-Design and the Non-Driving Related Task-Modality. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *61*(5), 839–849. https://doi.org/10.1177/0018720818814963

Roche, F., Thüring, M., & Trukenbrod, A. K. (2020). What happens when drivers of automated vehicles take over control in critical brake situations? *Accident Analysis & Prevention*, *144*. https://doi.org/10.1016/j.aap.2020.105588

Rodemerk, C., Habenicht, S., Weitzel, A., Winner, H., & Schmitt, T. (2012). Development of a general criticality criterion for the risk estimation of driving situations and its application to a maneuver-based lane change assistance system. *2012 IEEE Intelligent Vehicles Symposium*, 264–269. https://doi.org/10.1109/IVS.2012.6232129

Rohrmann, B. (1978). Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift Für Sozialpsychologie*, *9*, 222–245.

SAE International. (2018). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016)*. https://www.sae.org/standards/content/j3016_201806/

SAE International. (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016)*. https://www.sae.org/standards/content/j3016_202104/

Saito, T., Wada, T., & Sonoda, K. (2018). Control Authority Transfer Method for Automated-to-Manual Driving Via a Shared Authority Mode. *IEEE Transactions on Intelligent Vehicles*, *3*(2), 198–207. https://doi.org/10.1109/TIV.2018.2804167

Schömig, N., Hargutt, V., Neukum, A., Petermann-Stock, I., & Othersen, I. (2015). The Interaction Between Highly Automated Driving and the Development of Drowsiness. *Procedia Manufacturing*, *3*, 6652–6659. https://doi.org/10.1016/j.promfg.2015.11.005

Scott, J. J., & Gray, R. (2008). A Comparison of Tactile, Visual, and Auditory Warnings for Rear-End Collision Prevention in Simulated Driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(2), 264–275. https://doi.org/10.1518/001872008X250674

Shahid, A., Wilkinson, K., Marcu, S., & Shapiro, C. M. (2011). Karolinska Sleepiness Scale (KSS). In A. Shahid, K. Wilkinson, S. Marcu, & C. M. Shapiro (Eds.), *STOP, THAT and One Hundred Other Sleep Scales* (pp. 209–210). Springer New York. https://doi.org/10.1007/978-1-4419-9893-4_47

Siebert, F. W., Oehl, M., & Pfister, H.-R. (2014). The influence of time headway on subjective driver states in adaptive cruise control. *Transportation Research Part F: Traffic Psychology and Behaviour*, *25*, 65–73. https://doi.org/10.1016/j.trf.2014.05.005

Tesla, Inc. (2022). *Autopilot*. https://www.tesla.com/autopilot

Toyota Automated Driving. (2020). https://automatedtoyota.com/docs/10/1a0b46b2c789165fdf296ad6d74e796f2b57837f9/ATwhitepaper.pdf

Tscharn, R., Naujoks, F., & Neukum, A. (2018). The perceived criticality of different time headways is depending on velocity. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 1043–1052. https://doi.org/10.1016/j.trf.2018.08.001

United Nations Economic Commission for Europe. (2014). *Report of the sixty-eighth session of the Working Party on Road Traffic Safety*.

U.S. Department of Transportation, National Highway Traffic Safety Administration. (2013). *Preliminary Statement of Policy Concerning Automated Vehicles*.

van den Beukel, A. P., & van der Voort, M. C. (2013). The influence of time-criticality on Situation Awareness when retrieving human control after automated driving. *Proceedings of 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 2000–2005. https://doi.org/10.1109/ITSC.2013.6728523

Vogel, K. (2003). A comparison of headway and time to collision as safety indicators. *Accident Analysis & Prevention*, *35*(3), 427–433. https://doi.org/10.1016/S0001-4575(02)00022-2

Vogelpohl, T., Kühn, M., Hummel, T., & Vollrath, M. (2019). Asleep at the automated wheel—Sleepiness and fatigue during highly automated driving. *Accident Analysis & Prevention*, *126*, 70–84. https://doi.org/10.1016/j.aap.2018.03.013

VW. (2020). *Der lange Weg zur Autonomie*. https://www.volkswagenag.com/de/news/stories/2018/01/the-long-path-to-autonomy.html#

Wada, T., Sonoda, K., Okasaka, T., & Saito, T. (2016). Authority transfer method from automated to manual driving via haptic shared control. *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 002659–002664. https://doi.org/10.1109/SMC.2016.7844641

Walch, M., Lange, K., Baumann, M., & Weber, M. (2015). Autonomous driving: Investigating the feasibility of car-driver handover assistance. *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular*

Applications, 11–18. https://doi.org/10.1145/2799250.2799268

Winkler, S., Kazazi, J., & Vollrath, M. (2018). Practice makes better – Learning effects of driving with a multi-stage collision warning. *Accident Analysis & Prevention*. https://doi.org/10.1016/j.aap.2018.01.018

World Health Organization. (2021, June 21). *Road traffic injuries*. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Wörle, J., Metz, B., & Baumann, M. (2021). Sleep inertia in automated driving: Post-sleep take-over and driving performance. *Accident Analysis & Prevention*, *150*, 105918. https://doi.org/10.1016/j.aap.2020.105918

Yi, B., Cao, H., Song, X., Zhao, S., Guo, W., & Li, M. (2022). How to identify the take-over criticality in conditionally automated driving? An examination using drivers' physiological parameters and situational factors. *Transportation Research Part F: Traffic Psychology and Behaviour*, *85*, 161–178. https://doi.org/10.1016/j.trf.2021.12.007

Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accident Analysis & Prevention*, *78*, 212–221. https://doi.org/10.1016/j.aap.2015.02.023

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. *Accident Analysis & Prevention*, *92*, 230–239. https://doi.org/10.1016/j.aap.2016.04.002

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. *Transportation Research Part F: Traffic Psychology and Behaviour*, *50*, 65–79. https://doi.org/10.1016/j.trf.2017.07.008

Zhang, B., de Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *64*, 285–307. https://doi.org/10.1016/j.trf.2019.04.020

5 Appendix

A Table 2: Selection of research topics about system-initiated take-overs and corresponding studies

Research topics	**Driving studies**
Repeated experience of take-overs	Brandenburg and Roche (2020)
	Gold et al. (2015)
	Hergeth et al. (2015, 2017)
	Körber et al. (2015)
	Payre et al. (2016)
	Roche et al. (2018)
Time budget	Damböck et al. (2012)
	Gold et al. (2013)
	Mok, Johns, Lee, Ive, et al. (2015)
	Mok, Johns, Lee, Miller, et al. (2015)
	Roche and Brandenburg (2018, 2020)
	van den Beukel and van der Voort (2013)
	Walch et al. (2015)
	Yi et al. (2022)
Vehicle dynamics	Hu et al. (2019)
Take-over scenario	Damböck et al. (2012)
	Kerschbaum et al. (2015)
	Naujoks et al. (2014)
	Zeeb et al. (2017)
Traffic conditions	Gold et al. (2016)
	Radlmayr et al. (2014)
Non-driving related task	Feldhütter et al. (2017)
	Merat et al. (2012)
	Radlmayr et al. (2014)
	Roche et al. (2018)
	Schömig et al. (2015)
	Zeeb et al. (2016)

A Table 2: Selection of research topics about system-initiated take-overs and corresponding studies

	Zeeb et al. (2017)
Pre-take-over state	Agrawal and Peeta (2021)
	Du et al. (2020)
	Feldhütter et al. (2018)
	Vogelpohl et al. (2019)
	Wörle et al. (2021)
Design of take-over requests	Borojeni et al. (2016)
	Brandenburg and Roche (2020)
	Forster et al. (2017)
	Lorenz et al. (2014)
	Ma et al. (2021)
	Melcher et al. (2015)
	Naujoks et al. (2014)
	Petermeijer et al. (2017)
	Politis et al. (2015)
	Politis et al. (2018)
	Roche et al. (2018)

Note. This selection is exemplary and not final. Further research topics on system-initiated take-overs and more studies on each topic exist.

B Overview of included publications

Publication 1:

- Bibliographical information: Roche, F. (2021). Assessing subjective criticality of take-over situations after automated driving: Validation of two rating scales. *Accident Analysis & Prevention*, *159*, 1–13. https://doi.org/10.1016/j.aap.2021.106216
- Version: Publisher version
- Can be found in appendix C

Publication 2:

- Bibliographical information: Roche, F., Thüring, M., & Trukenbrod, A. K. (2020). What happens when drivers of automated vehicles take over control in critical brake situations? *Accident Analysis & Prevention*, *144*. https://doi.org/10.1016/j.aap.2020.105588
- Version: Publisher version
- Can be found in appendix D

Publication 3:

- Roche, F., Becker, S., & Thüring, M. (2022). What happens when drivers of automated vehicles take over control in critical lane change situations? *Transportation Research Part F: Traffic Psychology and Behaviour*, *84*, 407–422. https://doi.org/10.1016/j.trf.2021.11.021
- Version: Publisher version
- Can be found in appendix E

C Publication 1: Assessing the criticality of driving situations after automated driving: Validation of two scales

Accident Analysis and Prevention 159 (2021) 106216

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Assessing subjective criticality of take-over situations: Validation of two rating scales

Fabienne Roche

Technische Universität Berlin, Fachgebiet Kognitionspsychologie und Kognitive Ergonomie, MAR 3-2, Marchstraße 23, Berlin, 10587, Germany

ARTICLE INFO

Keywords:
Automated driving
Driver-vehicle interaction
Driver behavior
Criticality
Take-over behavior
Scale validation

ABSTRACT

Assessing subjective criticality of take-over situations is crucial for understanding of take-over behavior and comparing studies. However, no validated rating scales exist that assess subjective criticality of take-over situations. In a driving simulator study, two rating scales, the Scale of Criticality Assessment of driving situations from Neukum et al. (2008) and the Criticality Rating Scale, were tested on their validity to assess the subjective criticality of take-over situations. Besides, the subjective and behavioral changes over the repeated experience of take-over situations were investigated. Twenty-five participants experienced a set of five take-over situations with varying time-to-collisions (TTC) at the moment of the take-over request, twice. After each of the first five take-over situations, participants rated the criticality on one scale, after each of the second five situations on the other scale. Correlation coefficients between TTCs and criticality ratings for each scale were calculated. Also, the changes of subjective and behavioral measures over the trials were investigated. Correlation coefficients indicated a strong correlation between criticality ratings and TTCs. Hence, both scales are equally valid for the assessment of the criticality of take-over situations. The repeated experience of the take-over situations did not affect effort ratings, take-over times, or steering wheel positions. But brake input decreased with increasing practice, indicating a safer take-over behavior. Hence, results of studies with repeated experience of take-over situations are relatively valid as only brake behavior changed with increasing practice.

1. Introduction

In the past years, human factors researchers have accumulated an impressive amount of knowledge, especially on the take-over process (Gold et al., 2016; Jamson et al., 2013; Körber et al., 2016; Murata et al., 2013; Politis et al., 2014; Roche et al., 2018; SAE International, 2018). It was observed that different characteristics of take-over situations may heavily influence the take-over behavior and subjective experience (Damböck et al., 2012; Gold et al., 2016, 2013; Radlmayr et al., 2014; Roche and Brandenburg, 2020, 2018). One of these characteristics is the objective criticality of the take-over situation which is determined by situational parameters. For example, take-over situations with low time budgets are objectively more critical than situations with high time budgets. The objective criticality affects the take-over behavior and subjective criticality, e.g. more extreme behavior and higher subjective criticality when the situation is more critical. There are many options to assess take-over behavior, such as take-over times or steering behavior. It provides insights into how drivers behave depending on different situational parameters. In contrast, to our knowledge, no validated instrument exists to assess the subjective criticality, even though, it supports the interpretation of observed take-over behavior and may enable comparisons between different take-over situations. Therefore, in the present study, two rating scales are validated regarding their suitability to assess the subjective criticality of take-over situations. Besides, subjective and behavioral changes of the repeated experience of take-over situations are investigated.

1.1. *Objective criticality of driving situations*

The objective criticality of a driving situation is 'the accident risk' (Rodemerk et al., 2012, p. 1). Hence, a driving situation, in which a collision is inevitable, constitutes the highest possible objective criticality (Rodemerk et al., 2012). Especially in automated driving, the objective criticality of a driving situation is crucial since it influences the take-over behavior (Gold et al., 2013; Roche and Brandenburg, 2020, 2018; Zhang et al., 2019).

The objective criticality of a driving situation may be determined by situational parameters such as time budget (Junietz et al., 2017), traffic

E-mail address: fabienne.roche@freenet.de.

https://doi.org/10.1016/j.aap.2021.106216
Received 9 February 2021; Received in revised form 30 April 2021; Accepted 24 May 2021

density, or visibility. Lower time budgets, higher traffic densities, or poor visibility lead to a higher objective criticality. In this paper, we focus on time budget. It can be quantified by time-to-collision (TTC). TTC describes the available time until a vehicle would collide with a reference object (Vogel, 2003). A reference object may be a preceding vehicle or a system boundary, such as an obstacle on the road. Hence, shorter TTCs indicate a more critical situation. These more critical situations may emerge in case the automated driving system reaches its limits or in the case of driver-initiated take-overs (Roche et al., 2020).

TTC is known to influence take-over behavior. Numerous driving simulator studies varied the TTC in take-over situations. Lower TTCs, hence more critical take-over situations, were associated with lower take-over times (Gold et al., 2013; Roche and Brandenburg, 2018, 2020; Zhang et al., 2019), higher decelerations (Roche et al., 2020; Roche and Brandenburg, 2018, 2020), and larger steering wheel angles (Roche et al., 2020; Roche and Brandenburg, 2018, 2020). While lower take-over times are a desirable behavior, high decelerations and extreme steering are a threat to the drivers' safety for the following reasons: This behavior may result in (a) vehicle instability, (b) rear-end collisions with following vehicles, (c) collisions with vehicles on neighboring lanes or (d) lane departures. Indeed, more critical take-over situations in terms of lower TTCs led to higher error rates, such as collisions or missing lane changes (Damböck et al., 2012), more lane departures (Mok et al., 2015a; 2015b), and more collisions (Roche and Brandenburg, 2018). This impaired performance points at the threat of take-overs in situations with low TTCs.

1.2. *Subjective criticality of driving situations*

Analog to the definition of objective criticality, subjective criticality may be defined as the perceived threat or risk of a driving situation (Rodemerk et al., 2012). Hence, situations that are objectively more critical are highly likely to be perceived as more critical. However, next to the objective criticality, further aspects may affect the subjective criticality. These are individual parameters such as the driver's personality (Banet and Bellet, 2008; Mesken et al., 2007), fatigue (Feldhütter et al., 2018), familiarity with the take-over situation (Hergeth et al., 2017), or perceived capability (Fuller, 2011). For instance, it has been demonstrated that fatigued drivers rate the same situations as more critical than alert drivers indicating that they are more stressed (Feldhütter et al., 2018). And Hergeth et al. (2017) observed that criticality ratings decreased with increasing familiarity with the take-over situation.

There are two reasons why it is crucial to assess subjective criticality of take-over situations. First, situational parameters may have diverse effects on take-over behavior and may interact. Assessing the subjective criticality of take-over situations would promote the understanding of the observed behavior. Second, a criticality rating facilitates the comparability of driving situations and the evaluation of take-over behavior. Rodemerk et al. (2012) advocated for the need to compare driving situations and introduced a general criticality criterion to do so. Similarly, Jarosch and Bengler (2018) suggested a holistic view to evaluate the take-over behavior adequately rather than consider the parameters separately. They argue that, for example, a fast reaction cannot generally be evaluated as a good reaction, but has to be looked at in combination with steering and braking behavior. However, both criteria from Rodemerk et al. (2012) and Jarosch and Bengler (2018) are based on objective parameters, e.g. collision probability. They do not include the subjective aspect of a take-over situation. In contrast, Radlmayr et al. (2018) included a subjective rating parameter to evaluate take-over behavior. Together with two further parameters, it ought to promote the understanding and the comparability of take-over situations (Radlmayr et al., 2018). In line with Radlmayr et al. (2018), we argue that a validated and anchored scale to assess subjective criticality would enable the comparability of different driving situations and the evaluation of take-over behavior in driving simulator studies and real traffic.

In numerous studies, different scales for assessing the subjective criticality of driving situations have been employed. These are, for example, a multi-item Likert-scale (Banet and Bellet, 2008), an eleven-point, single item scale developed by Neukum et al. (2008), or seven-point Likert-scales (Radlmayr et al., 2018; Roche and Brandenburg, 2020, 2018). However, to our knowledge, no publication is available that validates one of them.

1.2.1. *Scale of criticality assessment of driving situations*

The *Scale of Criticality Assessment of driving situations* (SCA; Neukum et al., 2008) is a scale that assesses subjective criticality of a driving situation (see Fig. 1, left in English, right in German). It was already used in various studies (Hergeth et al., 2017; Naujoks et al., 2017; Neukum et al., 2008; Siebert et al., 2014). The scale is a modified version of the judgment *Scale for the Assessment of the Experienced Degree of Disturbance* (Neukum and Krüger, 2003) that was developed based on the Cooper-Harper-Scale (Cooper and Harper, 1969). It is an eleven-point, single item scale and based on a two-step rating procedure. First, participants are asked to rate the criticality of the driving situation by selecting one of the five verbal categories: 'imperceptible' (0 pt.), 'harmless' (1–3 pts.), 'unpleasant' (4–6 pts.), 'dangerous' (7–9 pts.), 'uncontrollable' (10 pts., translation based on Naujoks et al., 2017). Participants are instructed that the driving situation shall be rated concerning the necessary compensatory effort. Second, they are asked to specify their rating by selecting one of the three numerical subcategories of each category (right area in Fig. 1). It was assumed that 'imperceptible' and 'uncontrollable' are not divisible any further. Hence, these both extreme categories have only one subcategory, 0 respectively 10 pts.

This scale holds advantages and disadvantages. On the one hand, Neukum et al. (2008) state that an advantage of this scale is the threshold distinguishing between tolerable and intolerable situations (rating above 6). The magnitude of ratings should be, therefore, comparable between participants. Another advantage is its sensitivity that allows for differentiated ratings of subjective criticality across different driving situations (Neukum and Krüger, 2003). Besides, the original scale was highly accepted by naïve and expert participants (Neukum and Krüger, 2003). On the other hand, the scale does not take into account whether the driver's perceived capability to deal with the driving situation affects the rating, as suggested by Fuller (2011). Hence, drivers that feel very capable might rate a driving situation as tolerable, while others rate it as intolerable. In addition, it is questionable whether the differences between the numerical values are equidistant. For example, it may be assumed that the experienced differences within one category (e.g. two vs. three) differ from the experienced differences between two categories (e.g. three vs. four). Also, the effort for the application and analysis is high because the instruction takes longer and it has to be ensured that the rating of the verbal category corresponds to the rating of the numerical subcategory.

1.2.2. *Criticality rating scale*

The *Criticality Rating Scale* (CRS) is a modification of a criticality scale used in previous studies (Roche and Brandenburg, 2018, 2020). It was modified in the course of this research project aiming at compensating known disadvantages of existing rating scales. In the instruction of the item, it is clearly stated that the criticality should be rated concerning real driving situations. The wording of the question 'How critical did you perceive the experienced driving situation?' avoids complicated syntax, specific terms, and ambiguity as recommended by Moosbrugger and Kelava (2020). The Criticality Rating Scale consists of a single item rating scale. A continuous scale with tick marks is used to visualize the gradation of the rating (see Fig. 2). This is in contrast to the SCA, but similar to the NASA-TLX scale (Hart and Staveland, 1988). The NASA-TLX is a well-known and widely used tool for the assessment of perceived workload (Hart, 2006). In line with the NASA-TLX, the CRS is

Fig. 1. Scale of Criticality Assessment of driving situation (SCA) in English (left, depicted from Naujoks et al., 2017) and in German (right, Neukum et al., 2008). The German version was used in the present study.

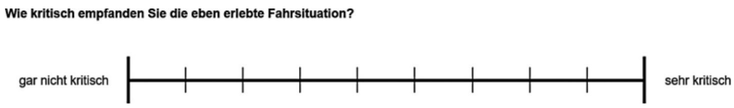

Fig. 2. The Criticality Rating Scale (CRS) with poles 'not critical at all' (1 pt.) and 'very critical' (100 pts.).

designed with 100 points (1−100). The poles are labeled with 'not critical at all' (1 pt., in German 'gar nicht kritisch') and 'very critical' (100 pts., in German 'sehr kritisch'). The verbal labeling is based on the results of Rohrmann's study on rating scales (Rohrmann, 1978). He found that the German versions of 'not at all' and 'very' were rated as the lowest and highest intensity terms with small scatter among eighteen terms. To allow the perception of an equidistant gradation, no more verbal or numerical anchors had been employed. Moosbrugger and Kelava (2020) recommended to omit a middle answer category because it is often used as a fallback option when the participant does not understand the question, refuses to answer, or does not know the answer. Hence, the number of ticks of the CRS was even to omit a middle answer category (see Fig. 2). In contrast, the SCA has a middle category (5 pts.). In doing so, the assumed disadvantages of the SCA shall be addressed.

The advantages of this scale are the following. First, less inter-individual interpretations are assumed since only the poles are labeled as opposed to the SCA. Thereby, a different understanding of the used labels by the raters is less likely. Second, the rating differences between all scale points should be equidistant since the scale points are not additionally labeled. Third, the test efficiency is supposed to be higher than with the SCA due to the short instruction and the familiar rating system. A potential disadvantage of the CRS is the missing threshold that would enable comparability of ratings between raters as assumed with the SCA.

1.3. Test validity

Tests should be valid to ensure they are truly measuring what they are supposed to measure (Hartig et al., 2008). The Frey (2018) state that test validity is the most important quality criteria of a test, next to objectivity and reliability. A test is objective when its result is independent of the experimenter and analyzer (Hartig et al., 2008). A test is reliable when it is precise, i.e. an elastic measuring tape would not be reliable when measuring length (Hartig et al., 2008). Objectivity and reliability are requirements to ensure test validity. It should be noted that none of the quality criteria is binary, hence, a test can be more or less valid, objective, or reliable. For the sake of this study, we assumed that both scales achieve a certain level of reliability due to their wording and scale design. Besides, we aimed at establishing objectivity when applying the scales (more details see 2 Method section).

Different types of test validity exist, among them construct validity. It describes the extent to which a test examines a psychological trait or construct, as defined by theory (Cronbach and Meehl, 1955). Construct validity is considered as the most fundamental type of test validity (Wainer and Braun, 2013) and composes of convergent and discriminant validity. Convergent validity is present when measurements of a construct that are recorded with different methods correlate strongly (Moosbrugger and Kelava, 2020). Usually, a new method is validated by means of another established method. Discriminant validity is high when measurements of different constructs that are recorded using the same or different methods correlate weakly (Moosbrugger and Kelava, 2020). In this study, we focus on investigating the convergent validity of both scales.

1.4. Repeated experience of take-over situations

The repeated experience of similar take-over situations is quite common in driving simulator studies due to the experimental setting and test efficiency. This may lead to subjective and behavioral changes. On the one hand, it would be plausible that take-over behavior deteriorates over the course of an experiment. Reasons for deterioration are increasing fatigue or increasing trust in the system (Hergeth et al., 2016). Indeed, there are indications that with increasing practice take-over behavior becomes riskier: maximal deceleration increases (Brandenburg and Roche, 2020) and observation decreases (Hergeth et al., 2016; Roche et al., 2018). On the other hand, it could be that take-over behavior improves due to the increasing practice. This was demonstrated by Hergeth et al. (2017), Körber et al. (2016), and Payre et al. (2016) concerning decreasing take-over times, larger TTCs, lower maximal lateral accelerations, and lower maximal longitudinal decelerations. This shows that the evidence regarding the effect of experience is not unambiguous.

1.5. Research questions

The two rating scales are supposed to measure the subjective criticality of take-over situations. Hence, testing their validity requires checking whether they properly measure criticality. As a first step, the convergent validity of the scales is tested. Therefore, a variation of objective criticality in the take-over situations should be represented in the criticality ratings. Hence, the following research questions (RQ) are investigated:

- Research question 1: Is the Scale of Criticality Assessment of driving situations (SCA) a valid tool for the assessment of subjective criticality in take-over situations?
- Research question 2: Is the Criticality Rating Scale (CRS) a valid tool for the assessment of subjective criticality in take-over situations?
- Research question 3: Do both scales differ regarding their validity?

For this, objective criticality is varied by the time-to-collision, an established method to vary criticality of take-over situations. In case the ratings and TTC-values correlate strongly, a high convergent validity can be assumed.

The repeated experience of similar take-over situations would indicate whether the rating scales are robust in assessing subjective criticality and whether increasing familiarity affects the ratings. In addition, since our participants experience several monotonous trials, it likely leads to fatigue and lower arousal. De Waard (2002) stated that passive fatigue may be compensated by increasing effort. Hence, increasing perceived effort ratings over the repetition of trials would demonstrate passive fatigue. Furthermore, the available research on the repeated experience of take-over situations indicates that behavioral change may take place. This leads to the fourth research question:

- Research question 4: Do drivers' criticality and effort ratings and take-over behavior change over the repeated experience of take-over situations?

2. Method

The objective criticality of the take-over situation is manipulated by the time-to-collision to a stationary obstacle at the moment of the take-over request. A lane change was chosen as the take-over situation because it is one of the most common maneuvers on the highway, where higher levels of automated driving will be deployed first (Bellem et al., 2017). In a driving simulator, participants experienced five take-over situations twice that varied regarding time-to-collisions. After each of the first five take-over situations, participants rated the criticality on one scale, after each of the second five situations on the other scale.

2.1. Participants

Twenty-five persons (13 women, 12 men) between 21 and 37 years of age ($M = 27.3$ years, $SD = 4.8$ years) took part in the driving simulator study. An a-priori power analysis with G*Power (Version 3.1.9.2) revealed that a sample size of 23 participants was required to detect a correlation of $r = .5$ with a given alpha of .05 and a power of .80. With 25 participants, each of the five TTC-values could be presented at each position across the experiment five times, e.g. five participants experienced the shortest TTC-value in the first trial, five in the second, and so on. All participants had to be German native or near-native speakers to follow the German instructions. They were required to have been holding a driving license for a minimum of two years ($M = 9.6$ years, $SD = 4.7$, $Max = 19$ years). Twenty participants (80 %) were students. Ten participants (40 %) reported having experience with advanced driver-assistance systems, such as adaptive cruise control or lane change assistance systems. On average, they used their car at 2.8 days per week. The student participants received course credits as gratification. The experiment was approved by the ethics committee of the Department of Psychology and Ergonomics of Technische Universität Berlin, Germany, and its conditions complied with the tenets of the Declaration of Helsinki. Participants gave their informed consent before the experiment started.

2.2. Materials

The experiment was conducted in a mid-fidelity driving simulator of the Department of Psychology and Ergonomics of Technische Universität Berlin. The same driving simulator was used in Roche and Brandenburg (2020, 2018). It consists of a Volkswagen™ vehicle mock-up including Fanatec pedals, a Fanatec steering wheel, a dashboard, a seat, and a gear shift. Since the simulated vehicle was automatic, the gear shift and clutch pedal were irrelevant for this study. OpenDS 4.5 was utilized to simulate the driving environment: a two-lane rural road including a crash barrier and other vehicles on both lanes. The driving scene was projected on a screen placed at a distance of 0.80 m to the vehicle mock-up. The size of the projection screen was 3m × 1.70m. An image resolution of 1920 × 1080 pixels and a frequency of 60 Hz were used. A rear-view mirror was embedded in the projection (see Fig. 3). A driving automation corresponding to SAE-level 3 (SAE International, 2018) was active once the experimental trial started. It could be deactivated by a steering wheel or brake pedal input by the driver. A take-over was detected when the steering wheel positions exceeded 0.14 % or the brake pedal position exceeded 0.1 % (Roche and Brandenburg, 2020, 2018). Driving noise and auditory signals were played back via two speakers behind the driver seat. An iPad with standard factory settings was used to administer all questionnaires via the online survey service SoSci-Survey version 3.1 (www.soscisurvey.de). That way, the scales were always presented in the same manner and the ratings could be given without the experimenter being able to see them diminishing the interviewer effect (Bogner and Landrock, 2016). Due to that, the application of the scales was objective to a certain extent.

Subjective criticality was assessed with the Scale of Criticality Assessment of driving situations (SCA; Neukum et al., 2008; see Fig. 1 and section 1.2.1) or the Criticality Rating Scale (CRS; see Fig. 2 and section 1.2.2), depending on the block.

The effort-subscale of the NASA-TLX was used to assess the perceived effort (Hart and Staveland, 1988). Participants were asked to answer the question' How hard did you have to work to accomplish your level of performance?'. The scale is a unipolar, single item with tick marks to visualize the gradation of the rating from 'low' (0 pts.) to 'high' (100 pts., see Fig. 4). We used the German translation by Sepehr (1988). It was applied to assess a further aspect of subjective experience and to investigate possible fatigue over the repetition of trials because driver fatigue may be compensated by higher effort (de Waard, 2002).

2.3. Procedure and experimental design

The experiment consisted of an instruction phase, a familiarization phase, a training phase, an experimental phase, and a final interview. In the instruction phase, participants were welcomed and instructed about the procedure of the experiment and the handling of personal data. Then, they were asked to read and sign the informed consent, and answer a demographic questionnaire, i.e. age and profession. The questionnaire and all following rating scales and questions were presented on the iPad.

In the familiarization phase, participants drove on a two-lane highway for about 3 min in the driving simulator. They practiced accelerating, decelerating, and lane changing to familiarize themselves with the driving simulator.

In the training phase, participants were introduced to the automated system and the driving task. The system was designed to take over longitudinal and lateral control for specific driving tasks, depicting a system at SAE-level 3 (SAE International, 2018). Each of the training trials started with an automatic acceleration of the participant's vehicle to 100 km/h. The vehicle drove on the right lane of a two-lane rural road. Participants were instructed to take hands off the steering wheel

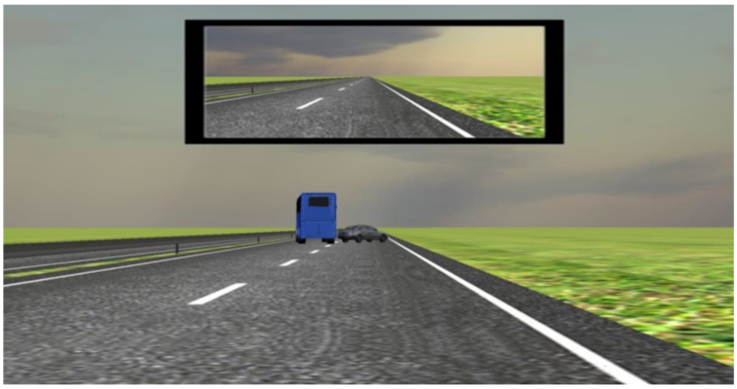

Fig. 3. Take-over situation with 2.5 s TTC, the crashed vehicle in grey (right), and the lead vehicle in blue (left). The rectangle in the upper-middle represents the rear-view mirror.

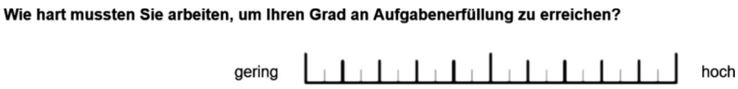

Fig. 4. Effort-subscale of the NASA-TLX in German (Hart and Staveland, 1988) ranging from 'low' to 'high'.

and feet off the pedals while driving automated. Upon an acoustic cue, they were instructed to take back control by steering or braking and steer around a construction work on their lane. The acoustic cue consisted of two consecutive sounds with a duration of 0.5 s each, a frequency of 780 Hz, and a volume of approximately 80 dB. Five training trials were driven with varying TTCs at the moment of the acoustic cue: 3.10, 3.35, 3.60, 3.85, and 4.10 s. The TTC-values were on a medium range and the order was balanced across participants. After the last training trial, participants rated the subjective criticality on the SCA and the CRS and their perceived effort on the subscale of the NASA-TLX. The ratings did not enter into the analysis; they were rather applied so that the participants got used to the rating scales.

In the experimental phase, the participants experienced two blocks of five experimental trials each. In each experimental trial, a lane change served as take-over situation with varying TTCs at the moment of the take-over request (TOR). In one block, participants rated the criticality on the SCA, in the other block, on the CRS. The sequence of blocks was balanced across participants, i.e. 13 participants started rating subjective criticality on the SCA, 12 participants on the CRS. Similar to the training trials, the simulated vehicle started in automated mode executing longitudinal and lateral control. The automation was designed to accelerate to 100 km/h, keep the speed for the course of the trial, and drive at the center of the right lane. In all trials, the participant's vehicle followed a lead vehicle, a blue coach (see Fig. 3), and maintained a constant distance of 1.8 s time headway, i.e. the time it will take the participant's vehicle to reach the position of the lead vehicle.

After about 1 min in each trial, the take-over situation took place. In this situation, the participants' lane was blocked due to a broken vehicle. As soon as the lead vehicle changed lanes to avoid a collision, the obstacle became visible to the participant. The automation was able to detect the obstacle and requested the driver to take over by an acoustic cue. The same acoustic cue as in the training phase was used. The timing of the lane change maneuver of the lead vehicle, hence the timing of the TOR, was varied within-subjects. This resulted in different TTCs concerning the obstacle at the moment of the TOR. Five equidistant TTC-values were realized: 2.5, 3.0, 3.5, 4.0, 4.5 s. They were presented in a balanced order across participants. This means, over all participants, each TTC-value was presented at each position five times. However, the sequence of TTC-values between the two blocks was held constant for each participant to avoid any sequence effects on criticality ratings. The interval of 0.5 s between the TTC-values was chosen because a pretest showed that 0.5 s was large enough to cover a certain range of TTCs without having too many trials. Participants were instructed to take back control as fast and safely as possible by steering or braking upon the TOR. After steering around the obstacle or braking to a complete stop in front of the obstacle, the simulation was switched off and the simulation was stopped. On the iPad, participants rated the subjective criticality and the perceived effort. An instruction trial was added before each block to avoid surprise effects on the criticality ratings (see Fig. 5). For the instruction trials, a TTC from the medium spectrum was used (3.63 s). This resulted in one instruction and five experimental trials per block, hence, participants experienced two instruction trails and ten experimental trials in total (see Fig. 5).

In a final interview, participants were asked for their personal preference regarding the two rating scales. Participants were debriefed of the scope of the experiment. Overall, the experiment lasted about 90 min.

2.4. Dependent variables

The subjective criticality rating was the main dependent variable. It was assessed with the SCA [0−10 pts.] or CRS [1−100 pts.] at the end of each trial. Besides, the effort-subscale of the NASA-TLX [0−100 pts.], take-over time [ms], maximal steering wheel position [%], and maximal brake pedal position [%] for each trial served as dependent variables. Take-over time was measured between the onset of the acoustic cue and the driver response in terms of the steering wheel or brake pedal input. For maximal steering wheel position and maximal brake pedal position, the highest values during the take-over were extracted per participant and trial.

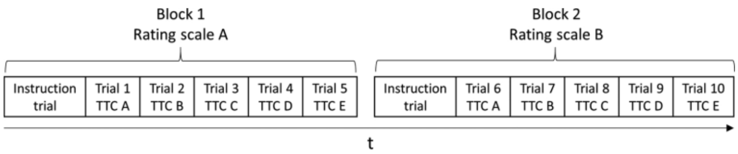

Fig. 5. Experimental course. The sequence of TTC-values (A-E) was balanced between-subjects and held constant between blocks (1 and 2). In the first block, criticality ratings were collected on rating scale A, in the second block, on rating scale B. The order of rating scales was balanced between-subjects.

2.5. Data analysis

For the analysis, R version 3.6.1 (R Core Team, 2019) was used. The correlation between TTC and the criticality ratings was calculated based on a method proposed by Bland and Altman (1995). This method accounts for repeated observations as in the present study. It is implemented in the R-package 'rmcorr' (Bakdash and Marusich, 2018). Degrees of freedom were calculated with Bakdash and Marusich's method available in rmcorr. In accordance to Hemphill (2003), a correlation coefficient $r < 0.21$ indicates a weak correlation, between 0.21 and 0.33 a medium correlation, and $r > 0.33$ a strong correlation. The correlation coefficients were used to answer the question of whether the two scales are valid tools for the assessment of criticality of take-over situations (RQ 1 and 2). Besides, it was tested whether the ratings of each TTC-value differed from the remaining ratings. Since we had paired samples and did not expect a normal distribution of the data, the Friedmann-test is used (Friedman, 1937). In case a significant difference was found, a post-hoc analysis was calculated using the Nemeyi-test (Nemenyi, 1962) of the R-package 'PMCMR' (Pohlert, 2014). The Bonferroni-method was used to adjust p-values.

To compare the two rating scales against each other (RQ 3), it was tested whether the correlation coefficients differed significantly based on a method suggested by Eid et al. (2017). The method determines a z-value of the two fisher-Z transformed correlation coefficients that can be tested on significance.

For the analysis of the change of ratings and behavior over trials (RQ 4), mixed-effects models for each dependent variable were calculated with the 'lme4'-package (Bates et al., 2015). The independent variable 'trial' served as a linear predictor. We accounted for inter-individual differences mentioned in the introduction and for the repeated measurement by adding a random intercept for each participant. Degrees of freedom were estimated with Satterthwaite's method available in the 'lmerTest'-package (Kuznetsova et al., 2017). The goodness-of-fit of each model is characterized by the marginal and conditional coefficient of determination (R^2, Nakagawa and Schielzeth, 2013).

3. Results

Since participants could rate subjective criticality either on the SCA or CRS, 125 trials (25 participants x 5 trials) were available for the analysis of the SCA- and 125 trials for the CRS-ratings. For the analysis of the perceived effort ratings and the behavioral data, data from ten trials per participant were available, resulting in 250 trials.

3.1. Correlation of the Scale of Criticality Assessment of driving situations with time-to-collision (RQ 1)

Fig. 6 visualizes the mean SCA-ratings for all five TTC-values. The mean SCA-rating across all TTC-values was 4.62 pts. ($SD = 1.9$). This mean value corresponds to the verbal category 'unpleasant' and is located at the threshold between tolerable and intolerable situations defined by Neukum et al. (2008). The ratings decrease with increasing TTC-values. The minimum ('imperceptible', 0 pts.) and maximum ('uncontrollable', 10 pts.) were not chosen by any participant.

The SCA-ratings correlated significantly with the TTC-values ($r_{TTC\_SCA}(99) = -0.59$, $p < .001$). Based on Hemphill (2003), the magnitude of the correlation coefficient indicates a strong correlation. The Friedman-test revealed that the SCA-ratings from at least two TTC-values differed significantly from each other ($\chi^2(4) = 35.36$, $p < .001$). The post-hoc test showed a significant difference between the most critical TTC-value (2.5 s) and the two least critical ones (4.0 s resp. 4.5 s, see the adjusted p-values for all comparisons in Table 1).

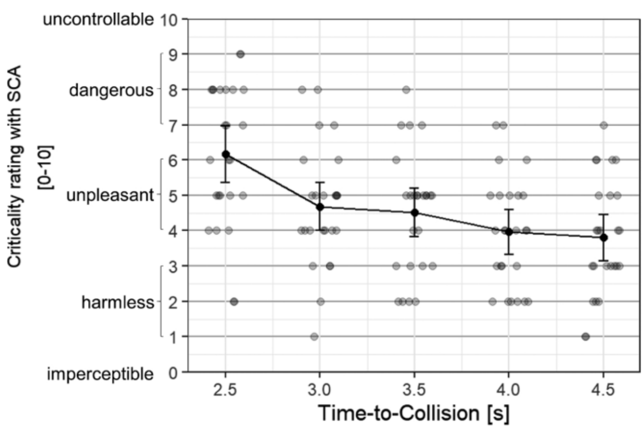

Fig. 6. Means, standard errors, and raw values of the Scale of Criticality Assessment of driving situations (SCA) for each time-to-collision-value from 2.5 to 4.5 s.

Table 1
Friedman-test results for the SCA-ratings. Adjusted p-values for post-hoc comparisons of SCA-ratings between all TTC-values.

Time-to-collision	2.5 s	3.0 s	3.5 s	4.0 s
3.0 s	.715			
3.5 s	.229	1		
4.0 s	< .001 ***	1	1	
4.5 s	< .001 ***	1	1	1

Significance symbols: *: p < .05, **: p < .01, ***: p < .001.

3.2. Correlation of the Criticality Rating Scale with time-to-collision (RQ 2)

The mean CRS-rating across all TTC-values was 45.98 pts. (SD = 25.2). Fig. 7 visualizes the means, standard errors, and raw values of the CRS. As expected, with increasing TTC, the CRS-ratings decreased (see Fig. 7). The minimum value (‘not critical at all’, 1 pt.) was selected 16 times, while the maximum (‘very critical’, 100 pts.) was never selected.

The correlation between CRS-ratings and TTC-values was highly significant ($r_{TTC\_CRS}(99) = -0.66$, $p < .001$). Again, this represents a strong correlation. The Friedman-test revealed that at least two CRS-ratings differed significantly from each other ($\chi^2(4) = 46.53$, $p < .001$). The post-hoc comparisons showed that in four cases the CRS-ratings differed from each other (see adjusted p-values in Table 2). These significant differences were between the most critical TTC-value (2.5 s) and the two less critical ones (4.0 s and 4.5 s), similar to the SCA-ratings. Also, the CRS-ratings of the TTC-value 3.0 s differed significantly from 4.0 s and 4.5 s.

3.3. Comparison of both scales (RQ 3)

For the comparison, the two correlation coefficients of the SCA and the CRS were used. The method suggested by Eid et al. (2017) revealed that the coefficients of the scales did not differ significantly ($z = 0.52$, $p = .603$).

The final interview showed that 84 % of the participants ($N = 21$) preferred the SCA for assessing subjective criticality of a take-over situation. 42.3 % of the participants ($N = 9$) reasoned their voting with the subdivision of the SCA into verbal and numerical categories. 38.1 % of the participants ($N = 8$) preferred the SCA due to the better description of the take-over situation by the verbal categories. Two of the participants preferring the CRS stated that the labeling of the poles were better suited for the take-over situation.

3.4. Repeated experience of take-over situations (RQ 4)

Mixed-effect models were calculated to investigate the change of ratings and take-over behavior over the repeated experience of the experimental trials. Significant estimates of the factor ‘trial’ would indicate a change of ratings or take-over behavior over the repeated experience. The statistical results are presented in Table 3 and the mean-values and standard errors per trial are plotted in Fig. 8. Only the maximal brake pedal position decreased significantly over trials. Descriptively, the SCA-, CRS-, and perceived effort ratings decreased slightly over the trials (see negative estimates for trial in Table 3 and Fig. 8). However, the models showed that none of these changes reached significance (all $t < 2$, $p > .05$). Hence, the ratings, take-over times, and steering wheel positions were not affected by the repeated experience of take-over situations. For all models, the marginal coefficient of determination was very small, hence, the variance explained by the fixed factor ‘trial’ was very low (below 1 %).

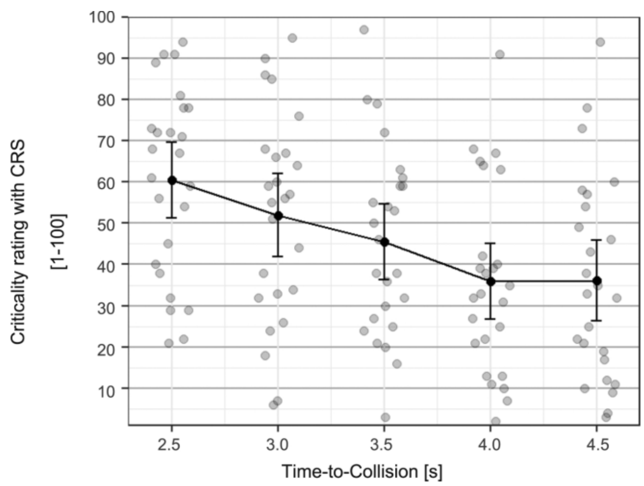

Fig. 7. Means, standard errors, and raw values of the Criticality Rating Scale (CRS) for each time-to-collision-value from 2.5 to 4.5 s.

Table 2
Friedman-test results for the CRS-ratings. Adjusted p-values for post-hoc comparisons of CRS-ratings between all TTC-values.

Time-to-collision	2.5 s	3.0 s	3.5 s	4.0 s
3.0 s	1			
3.5 s	.229	1		
4.0 s	< .001 ***	.009 **	.636	
4.5 s	< .001 ***	.014 *	1	1

Significance symbols: *: $p < .05$, **: $p < 0.01$, ***: $p < .001$.

Table 3
Summary of statistics for the repeated experience of trials of all dependent variables.

Criticality rating on SCA [0-10]	Estimate	Std. Error	df	t-value	p-value
Intercept	4.98	0.39	79.65	12.86	< .001 ***
Trial	-0.12	0.09	100	-1.32	0.191
Variance explained: $R^2_{marginal}$ = 0.8 %, $R^2_{conditional}$ = 41.9 %; N_{trials} = 125					

Criticality rating on CRS [1-100]	Estimate	Std. Error	df	t-value	p-value
Intercept	49.02	5.09	51.97	9.62	< .001 ***
Trial	-1.02	0.97	100	-1.05	.297
Variance explained: $R^2_{marginal}$ = 0.3 %, $R^2_{conditional}$ = 62.6 %; N_{trials} = 125					

Perceived effort rating [0-100]	Estimate	Std. Error	df	t-value	p-value
Intercept	41.41	4.05	48.88	10.21	< .001 ***
Trial	-0.41	0.40	225	-1.04	.299
Variance explained: $R^2_{marginal}$ = 0.2 %, $R^2_{conditional}$ = 44.3 %; N_{trials} = 250					

Take-over time [ms]	Estimate	Std. Error	df	t-value	p-value
Intercept	764.27	32.34	43.13	23.63	< .001 ***
Trial	-3.37	2.90	225	-1.16	.246
Variance explained: $R^2_{marstginal}$ = 0.3 %, $R^2_{conditional}$ = 51.2 %; N_{trials} = 250					

Maximal steering wheel position [%]	Estimate	Std. Error	df	t-value	p-value
Intercept	0.94	0.06	146.07	16.46	< .001 ***
Trial	0.00	0.01	225	0.24	.809
Variance explained: $R^2_{marginal}$ = 0.0 %, $R^2_{conditional}$ = 9.1 %; N_{trials} = 250					

Maximal brake pedal position [%]	Estimate	Std. Error	df	t-value	p-value
Intercept	57.43	4.88	39.79	11.76	< .001 ***
Trial	-0.81	0.41	225	-1.99	.048 *
Variance explained: $R^2_{marginal}$ = 0.7 %, $R^2_{conditional}$ = 56.4 %; N_{trials} = 250					

Significance symbols: *: $p < .05$, **: $p < .01$, ***: $p < .001$.

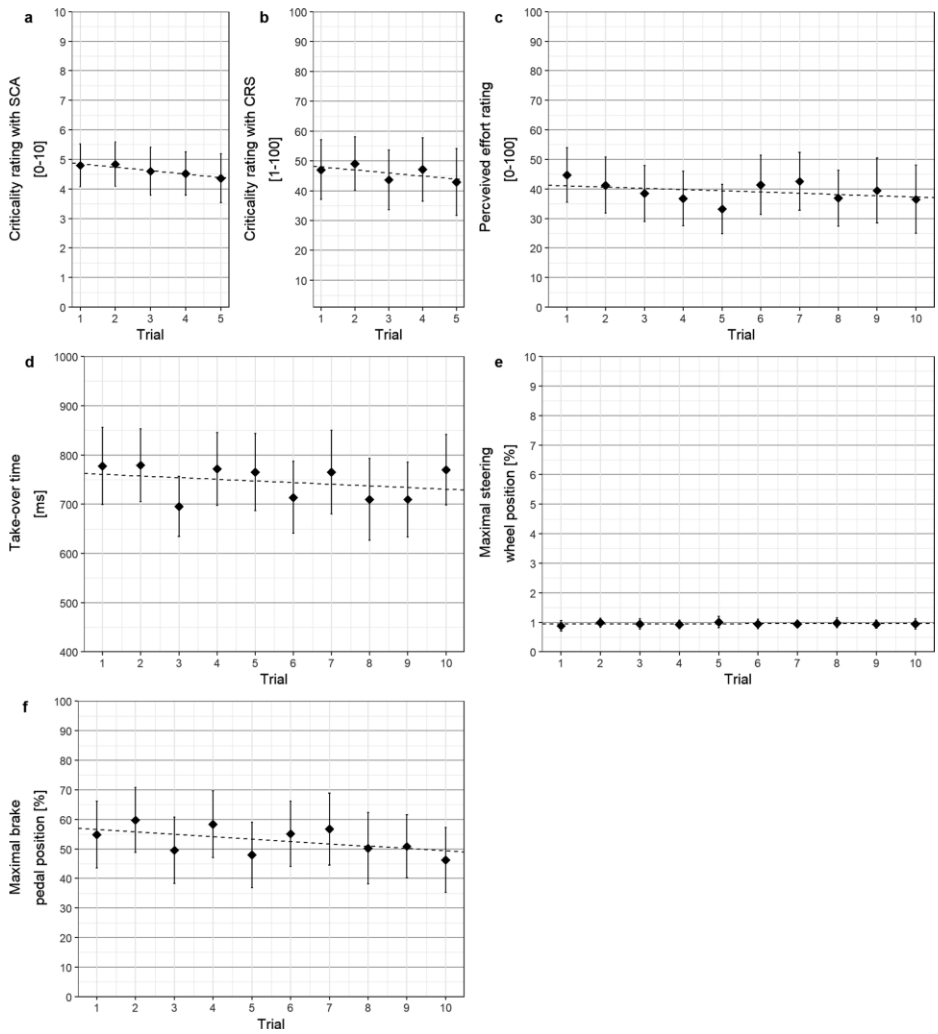

Fig. 8. Means, standard errors, and regression lines of the dependent variables per trial over the repeated experience of take-over situations.
Note. Five trials were evaluated on the SCA, five with the CRS. Concerning the remaining dependent variables, values are available for all ten experimental trials.

4. Discussion

The present study investigated whether two rating scales, the Scale of Criticality Assessment of driving situations (SCA) and the Criticality Rating Scale (CRS), are valid tools for the assessment of subjective criticality of take-over situations (RQ 1 and 2) and whether one is superior to the other one (RQ 3). Besides, the effects of the repeated experience of take-over situations on ratings and take-over behavior (RQ 4) were investigated. Participants experienced five experimental take-over situations twice that differed regarding time-to-collision. They provided their criticality rating either on the SCA or the CRS. Perceived effort ratings and take-over behavior were recorded in the ten experimental trials.

Before discussing the research questions, it should be noted that the take-over times were very small. This could be due to several reasons. First and in contrast to most other studies, our participants did not perform a non-driving related task. Hence, they could focus on the driving situations. Second, it could be that participants were highly trained to take over very fast after the training trials. Third, we assume that they were highly alert to expect a take-over by its frequent occurrence. Forth, the time budget used in this study was smaller than in most other studies on take-over time (2.5 s – 4.5 s in our study vs. 5 s and 7 s in Gold et al. (2013) or 8.6 s in Roche et al. (2018)). Previous research showed that shorter time budgets lead to shorter take-over times.

4.1. Research question 1 and 2: Are the Scale of Criticality Assessment of driving situations and the Criticality Rating Scale valid tools for the assessment of subjective criticality in take-over situations?

The study showed that both scales correlate strongly with the TTC-values that were varied to manipulate objective criticality of the take-over situations. This indicates that the SCA and the CRS are valid tools to assess the subjective criticality. The study paved the way of validating criticality rating scales in driving studies. However, convergent validity was tested, while different types such as discriminant or criterion validity were not investigated. This should be addressed in future studies. Furthermore, as mentioned in the introduction, validity is continuous and cut-off values for correlations for validity testing do not exist. Hence, one could argue that higher correlations are requested to infer validity. Besides, the conclusion is limited to a lane change and take-over situations in which objective criticality is varied by TTC. It is questionable whether the correlations between objective criticality and criticality ratings would be equally high in other maneuvers or when objective criticality is varied by different variables. For example, it could be that increasing traffic density from low to medium traffic would have a different effect on criticality ratings than an increase from medium to high traffic. Also, the rating scales are only validated for time-to-collisions in take-over situations. A transferability to driving situations in general is not given. Finally, it should be noted that the scales measure a general perception of criticality of take-over situations. Specific aspects such as collision risk or vehicle stability cannot be extracted. For this purpose, more comprehensive questionnaires would be necessary. Hence, future studies should validate the rating scales in different maneuvers and with other situational parameters, e.g. traffic density, to manipulate objective criticality and test different types of validity.

4.2. Research question 3: Do both scales differ regarding their validity?

Even though the two rating scales use different scale designs, the comparison of the correlation coefficients demonstrated that they do not differ regarding evaluation. Hence, the two scales are equally well suited for the assessment of subjective criticality in this specific take-over situation with this manipulation of objective criticality.

The results are noteworthy. On the one hand, the SCA is more time-consuming regarding instruction and processing than the CRS. More effort has to be put in explaining this scale since it is an unusual design. When processing the SCA, it has to be checked whether the rating of the first step (verbal category) corresponds to the rating of the second step (numerical subcategory). Besides, the correlation coefficient of the CRS was slightly higher and the CRS could better discriminate between different TTC-values as indicated by the higher amount of significant differences of the post-hoc comparisons. On the other hand, more participants preferred the SCA when rating subjective criticality of a take-over situation. Furthermore, as stated by Neukum et al. (2008), an advantage of the SCA is the threshold between tolerable and intolerable situations that is supposed to make ratings more comparable between raters. However, this reason was not yet proven. These aspects should make an impact on the researchers' decision on which scale to use in the future.

Apart from the research question, two additional insights concerning the two rating scales should be mentioned: First, the criticality ratings of the SCA and the CRS showed that differences of objective criticality are rather experienced with the more critical TTCs than with the less critical TTCs (see Figs. 6 and 7). This is in line with Siebert et al. (2014), who found rating differences between more critical THWs and no differences between less critical THWs to a lead vehicle. Siebert et al. (2014) interpreted this result as a threshold effect for the relation between objective criticality and subjective variables. While they used a car-following scenario, we likely observed the same effect in a different driving situation.

Second, participants neither used the minimum category of the SCA ('imperceptible') nor the maximum categories of the SCA ('uncontrollable') or the CRS ('very critical'). It seems as the lane change could not be ignored because no participant selected the minimum category. Concerning the maximum categories, the impression arises as none of the TTC-values was small enough to not be coped with because none of them was rated as maximal critical. It could be that the realization of the take-over situation did not achieve to cover a wide range of objective criticality. Hence, future studies may seek to cover a broader range of criticality.

4.3. Research question 4: Do drivers' criticality and effort ratings and take-over behavior change over the repeated experience of take-over situations?

Neither the ratings nor the take-over behavior changed over the repeated experience of the ten take-over situations, except maximal brake pedal position (RQ 4). This might be due to two reasons. First, the behavior and subjective experience likely changed within the five training trials. Hence, participants were already highly trained and habituated when the experiment started. Second, it could also be that subjective experience and behavior change within the first experimental trials and does not change in the following. Forster et al. (2019) found stabilized reaction times after three trials for transitions between SAE-level 2 and 3. Our analysis across all ten trials might have overruled a potential effect. In the present study, participants experienced twelve take-over situations, while participants of other studies experienced fewer situations, for example two in Hergeth et al. (2016) or six in Roche et al. (2018).

The criticality ratings of the take-over situations seem to be robust to a certain extent with the exception that the respondents possibly were already habituated to the take-over situations. This is a promising finding, as it suggests that even after the repeated experience of a take-over situation, the criticality ratings on the SCA and the CRS are still valid and comparable to the first rating.

Regarding the perceived effort ratings, the results showed that participants did not experience increased or decreased effort over trials, even though, the setting was monotonous with ten similar experimental and two instruction trials. Based on de Waard (2002), increasing fatigue due to the monotonous experimental setting may become apparent by increasing effort to cope with the situation. A reason why no change of perceived effort over trials was observed is that participants' increasing practice had compensated for the increasing passive fatigue. In consequence, participants might not have experienced increasing effort.

In line with Brandenburg and Roche (2020), we neither observed a change of take-over times nor of steering wheel positions over trials. The missing effects might be due to three reasons. First, similar to the perceived effort ratings, participants' practice likely increased due to the repeated experience of the ten take-over situations. This would allow drivers to anticipate future states and, usually, improve their performance (Endsley, 1995). Passive fatigue possibly increased at the same time. Hence, increasing practice and increasing fatigue might have compensated each other and led to a constant level of behavior. Second, as indicated earlier, take-over times and steering behavior might have changed within the training (and first experimental) trials and stabilized in the following. Such way, a significant change during the experimental trials was not detectable. Third, it could be that we observed a floor effect concerning take-over times because they were very small, making a faster reaction nearly impossible. Similarly, Brandenburg and Roche (2020) argued that a reason for the missing effect of repeated experience on take-over times might be a floor effect due to the very fast take-overs.

A decrease of brake pedal position was observed but no other behavioral change over the repeated experience. In contrast to our results, Brandenburg and Roche (2020) showed an increase in deceleration. It seems as brake behavior does not adapt as fast as other behavioral or subjective measures.

To conclude, these results indicate that studies with repeated experience of take-over situations are relatively valid as only brake behavior changed with increasing practice. However, it could be that subjective experience and behavior already adopted within the training or first experimental trials. Besides, it should be noted that the marginal coefficients of determination of all mixed-effect models were very small while the conditional coefficients of determination were quite high. This means that the fixed factor 'trial' did not explain much variance but the random intercept 'participant' did. Hence, the ratings and take-over behavior were mainly affected by inter-individual differences to rate or react rather than by the repeated experience of take-over situations.

4.4. Limitations

The study has some limitations that should be kept in mind when interpreting the results. First, the investigated take-over scenario was limited to one scenario: a lane change due to an obstacle in the participant's lane. Hence, the two rating scales have only been validated for this scenario. Besides, objective criticality was varied by manipulating TTC at the moment of the take-over request. Other characteristics of a take-over situation may also affect objective criticality. It should be tested whether similar correlations would be found if another take-over situation was used or if objective criticality would have been varied by different parameters.

Second, it must be noted that the driving simulator was mid-fidelity. The degree of immersion of the presented scenarios may be low compared to a high-fidelity simulator and the effect on perception and take-over behavior might differ from the one in real traffic. However, driving simulators allow low-cost and low-risk experiments in a controlled environment, especially for preliminary research (van Nes et al., 2010).

Third, due to the restrictions of the driving simulator, the lowest feasible TTC was 2.5 s and, due to the experimental design, the highest TTC was 4.5 s. However, the ratings show that almost the whole ranges of both scales were used, except the maximum and minimum categories. Future studies may aim at covering the whole range of the scales by presenting more and less objectively critical driving situations.

Forth, participants experienced many take-over situations in a row. This is an unrealistically high occurrence. Future studies should have a lower portion of take-over situations per session or more filler trials.

Fifth, the NASA-TLX was used to assess the development of passive fatigue over the course of the experiment. Rating scales on fatigue, e.g. Karolinska sleepiness scale (Shahid et al., 2011), would have been more appropriate.

Sixth, with a mean age of 27 years, the participants of the present study can be assigned to the younger population. Potential effects that come along with aging are impaired information processing (Salthouse, 1991) and increased hazard perception times (Horswill et al., 2008) which may result in slower take-overs. However, Körber et al. (2016) observed that take-over times did not differ between younger ($\leq$ 28 years) and older drivers ($\geq$ 60 years). But the older participants showed different take-over behavior than the younger ones, i.e. more and stronger braking and higher TTCs (Körber et al., 2016). Hence, it may be assumed that the study observed the best possible take-over behavior because mainly young drivers participated. Older participants might have shown different take-over behavior, i.e. larger take-over times, stronger braking, stronger steering.

5. Conclusion

The Scale of Criticality Assessment of driving situations (SCA) and the Criticality Rating Scale (CRS) are equally valid tools for the assessment of the subjective criticality of take-over situations. The ratings are robust over time. However, it should be noted that the two scales were only tested on convergent validity in this specific take-over situation of a lane change with this specific variation of objective criticality. Validation tests in other take-over situations and with different variations of objective criticality are pending. Besides, different types of validity should be investigated.

A behavioral change over the repeated experience of experimental take-over situations was only observed regarding braking. Possibly, subjective experience and take-over behavior adopted within the training trials, hence a change was not quantifiable. Effort ratings, take-over times, and steering wheel positions did not change.

Funding sources

This work has resulted from the interdisciplinary research project "Analysis and Support of Driver Interventions in Dynamic: Critical Situations during Highly Automated Driving" funded by the German Research Community (Deutsche Forschungsgemeinschaft, DFG, Grant No. 326727090).

Author statement

Roche. Assessing subjective criticality of take-over situations after automated driving: Validation of two rating scales

Conceptualization: Fabienne Roche
Data curation: Fabienne Roche and Oliver Blum
Formal analysis: Fabienne Roche
Funding acquisition: Fabienne Roche and Manfred Thüring
Investigation: Fabienne Roche and Oliver Blum
Methodology: Fabienne Roche and Oliver Blum
Project administration: Fabienne Roche and Manfred Thüring
Resources: Fabienne Roche and Manfred Thüring
Software: Fabienne Roche and Oliver Blum
Supervision: Fabienne Roche and Manfred Thüring
Validation: Fabienne Roche
Visualization: Fabienne Roche
Writing - original draft: Fabienne Roche
Writing - review & editing: Fabienne Roche, Manfred Thüring, and Stefan Brandenburg

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

We thank Oliver Blum for the cooperation regarding study planning and his careful data collection. We also thank Mario Lasch for his technical support and Manfred Thüring and Stefan Brandenburg for proof reading.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.aap.2021.106216.

References

Bakdash, J.Z., Marusich, L.R., 2018. rmcorr: Repeated Measures Correlation. https://CRAN.R-project.org/package=rmcorr.

Banet, A., Bellet, T., 2008. Risk awareness and criticality assessment of driving situations: a comparative study between motorcyclists and car drivers. IET Intell. Transp. Syst. 2 (4), 241. https://doi.org/10.1049/iet-its:20080037.

Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48.

Bellem, H., Klüver, M., Schrauf, M., Schöner, H.-P., Hecht, H., Krems, J.F., 2017. Can we study autonomous driving comfort in moving-base driving simulators? A validation study. Hum. Factors J. Hum. Factors Ergon. Soc. 59 (3), 442–456. https://doi.org/10.1177/0018720816682647.

Bland, J.M., Altman, D.G., 1995. Statistics notes: calculating correlation coefficients with repeated observations: part 1—correlation within subjects. BMJ 310 (6977), 446. https://doi.org/10.1136/bmj.310.6977.446.

Bogner, K., Landrock, U., 2016. Response Biases in Standardised SurveysResponse Biases in Standardised Surveys. GESIS Survey Guidelines. https://doi.org/10.15465/GESIS-SG_EN_016.
Brandenburg, S., Roche, F., 2020. Behavioral changes to repeated takeovers in automated driving: the drivers' ability to transfer knowledge and the effects of takeover request process. Transport. Res. Part F: Psychol. Behav. https://doi.org/10.1016/j.trf.2020.06.002.
Cooper, G.E., Harper, P., 1969. The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities, p. 60.
Cronbach, L.J., Meehl, P.E., 1955. Construct validity in psychological tests. Psychol. Bull. 52 (4), 281–302. https://doi.org/10.1037/h0040957.
Damböck, D., Farid, M., Tönert, L., Bengler, K., 2012. Übernahmezeiten beim hochautomatisierten fahren. Tagung Fahrerassistenz 16–28.
de Waard, D., 2002. Mental Workload. In Human Factors for Highway Engineers, 1st ed. Bingley, pp. 161–175 https://trid.trb.org/view/708725.
Eid, M., Gollwitzer, M., Schmitt, M., 2017. Statistik und Forschungsmethoden (5.). Beltz.
Endsley, M.R., 1995. Toward a theory of situation awareness in dynamic systems. Hum. Factors J. Hum. Factors Ergon. Soc. 37 (1), 32–64.
Feldhütter, A., Kroll, D., Bengler, K., 2018. Wake Up and Take Over! The effect of fatigue on the take-over performance in conditionally automated driving. Proceedings of 2018 21st International Conference on Intelligent Transportation Systems (ITSC) 2080–2085. https://doi.org/10.1109/ITSC.2018.8569545.
Forster, Y., Hergeth, S., Naujoks, F., Beggiato, M., Krems, J.F., Keinath, A., 2019. Learning to use automation: behavioral changes in interaction with automated driving systems. Transp. Res. Part F Traffic Psychol. Behav. 62, 599–614. https://doi.org/10.1016/j.trf.2019.02.013.
Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Stat. Assoc. 32 (200), 675–701. https://doi.org/10.1080/01621459.1937.10503522.
Fuller, R., 2011. Driver control theory. From task difficulty homeostasis to risk allostasis. Handbook of Traffic Psychology. Academic Press. https://doi.org/10.1016/B978-0-12-381984-0.10002-5.
Gold, C., Damböck, D., Lorenz, L., Bengler, K., 2013. Take over!" How long does it take to get the driver back into the loop?. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57, pp. 1938–1942. https://doi.org/10.1177/1541931213571433.
Gold, C., Körber, M., Lechner, D., Bengler, K., 2016. Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density. Hum. Factors 58 (4), 642–652. https://doi.org/10.1177/0018720816634226.
Hart, S.G., 2006. NASA-task load index (NASA-TLX); 20 years later. In: Th ANNUAL MEETING, 50, pp. 904–908. https://doi.org/10.1177/154193120605000909, 9.
Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task load index): results of empirical and theoretical research. Adv. Psychol. 52, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9.
Hartig, J., Frey, A., Jude, N., 2008. Validität. In: Moosbrugger, H., Kelava, A. (Eds.), Testtheorie und Fragebogenkonstruktion. Springer.
Hemphill, J.F., 2003. Interpreting the magnitudes of correlation coefficients. Am. Psychol. 58 (1), 78–79. https://doi.org/10.1037/0003-066X.58.1.78.
Hergeth, S., Lorenz, L., Vilimek, R., Krems, J.F., 2016. Keep Your Scanners Peeled: Gaze Behavior as a Measure of Automation Trust During Highly Automated Driving. Hum. Factors J. Hum. Factors Ergon. Soc. 58 (3), 509–519. https://doi.org/10.1177/0018720815625744.
Hergeth, S., Lorenz, L., Krems, J.F., 2017. Prior familiarization with takeover requests affects drivers' takeover performance and automation trust. Hum. Factors J. Hum. Factors Ergon. Soc. 59 (3), 457–470. https://doi.org/10.1177/0018720816678714.
Horswill, M.S., Marrington, S.A., McCullough, C.M., Wood, J., Pachana, N.A., McWilliam, J., Raikos, M.K., 2008. The hazard perception ability of older drivers. J. Gerontol. B Psychol. Sci. Soc. Sci. 63 (4), P212–P218. https://doi.org/10.1093/geronb/63.4.P212.
Jamson, A.H., Merat, N., Carsten, O.M.J., Lai, F.C.H., 2013. Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. Transp. Res. Part C Emerg. Technol. 30, 116–125. https://doi.org/10.1016/j.trc.2013.02.008.
Jarosch, O., Bengler, K., 2018. Rating of take-over performance in conditionally automated driving using an expert-rating system. In: Stanton, N. (Ed.), Advances in Human Aspects of Transportation (Vol. 786, Pp. 283–294). Springer International Publishing. https://doi.org/10.1007/978-3-319-93885-1_26.
Junietz, P., Schneider, J., Winner, H., 2017. Metrik zur Bewertung der Kritikalität von Verkehrssituationen und - szenarien, p. 12.
Körber, M., Gold, C., Lechner, D., Bengler, K., 2016. The influence of age on the take-over of vehicle control in highly automated driving. Transp. Res. Part F Traffic Psychol. Behav. 39, 19–32. https://doi.org/10.1016/j.trf.2016.03.002.
Kuznetsova, A., Bruun Brockhoff, P., Haub Bojesen Christensen, R., 2017. lmerTest: Tests in Linear Mixed Effects Models. https://CRAN.R-project.org/package=lmerTest.
Mesken, J., Hagenzieker, M.P., Rothengatter, T., de Waard, D., 2007. Frequency, determinants, and consequences of different drivers' emotions: an on-the-road study using self-reports, (observed) behaviour, and physiology. Transp. Res. Part F Traffic Psychol. Behav. 10 (6), 458–475. https://doi.org/10.1016/j.trf.2007.05.001.
Mok, B.K.-J., Johns, M., Lee, K.J., Ive, H.P., Miller, D., Ju, W., 2015a. Timing of unstructured transitions of control in automated driving. Intelligent Vehicles Symposium (IV) 1167–1172. https://doi.org/10.1109/IVS.2015.7225841.
Mok, B.K.-J., Johns, M., Lee, K.J., Miller, D., Sirkin, D., Ive, P., Ju, W., 2015b. Emergency, automation off: unstructured transition timing for distracted drivers of automated vehicles. Proceedings of the IEEE 18th Intelligent Transportation Systems (ITSC) 2458–2464. https://doi.org/10.1109/ITSC.2015.396.
Moosbrugger, Kelava, A., 2020. Testtheorie und Fragebogenkonstruktion. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-61532-4.
Murata, A., Kanbayashi, M., Hayami, T., 2013. Effectiveness of automotive warning system presented with multiple sensory modalities. In: Duffy, V.G. (Ed.), Digital Human Modeling and Applications in Health, Safety, Ergonomics, and Risk Management. Healthcare and Safety of the Environment and Transport (Vol. 8025, Pp. 88–97). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39173-6_11.
Nakagawa, S., Schielzeth, H., 2013. A general and simple method for obtaining *R* 2 from generalized linear mixed-effects models. Methods Ecol. Evol. 4 (2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x.
Naujoks, F., Purucker, C., Wiedemann, K., Neukum, A., Wolter, S., Steiger, R., 2017. Driving performance at lateral system limits during partially automated driving. Accid. Anal. Prev. 108, 147–162. https://doi.org/10.1016/j.aap.2017.08.027.
Nemenyi, P., 1962. Distribution-free multiple comparisons. J. Int. Biomet. Soc. 18 (2), 263.
Neukum, A., Krüger, H.-P., 2003. Fahrerreaktionen bei Lenksystemstörungen – Untersuchungsmethodik und Bewertungskriterien. Reifen - Fahrwerk - Fahrbahn. 1791, 297–318.
Neukum, A., Lübbeke, T., Krüger, H.P., Mayser, C., Steinle, J., 2008. ACC-Stop&Go: fahrerverhalten an funktionalen Systemgrenzen. Workshop Fahrerassistenzsysteme-FAS 141–150.
Payre, W., Cestac, J., Delhomme, P., 2016. Fully automated driving: impact of trust and practice on manual control recovery. Hum. Factors J. Hum. Factors Ergon. Soc. 58 (2), 229–241. https://doi.org/10.1177/0018720815612319.
Pohlert, T., 2014. The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR) [R]. https://CRAN.R-project.org/package=PMCMR.
Politis, I., Brewster, S., Pollick, F., 2014. Evaluating multimodal driver displays under varying situational urgency. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 4067–4076. https://doi.org/10.1145/2556288.2556988.
R Core Team, 2019. R: a Language and Environment for Statistical Computing (3.6.1) [Computer Software]. R Foundation for Statistical Computing. https://www.R-project.org/.
Radlmayr, J., Gold, C., Lorenz, L., Farid, M., Bengler, K., 2014. How traffic situations and Non-driving related tasks affect the take-over quality in highly automated driving. In: Proceedings of the 58th Human Factors and Ergonomics Society Annual Meeting, 58, pp. 2063–2067. https://doi.org/10.1177/1541931214581434.
Radlmayr, J., Ratter, M., Feldhütter, A., Körber, M., Prasch, L., Schmidtler, J., Yang, Y., Bengler, K., 2018. Take-overs in level 3 automated driving – proposal of the take-Over performance score (TOPS). In: Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T., Fujita, Y. (Eds.), Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018) (Vol. 823, Pp. 436–446). Springer International Publishing. https://doi.org/10.1007/978-3-319-96074-6_46.
Roche, F., Brandenburg, S., 2018. Should the urgency of an auditory-tactile takeover request match the situational criticality? Proceedings 2018 21st International Conference on Intelligent Transportation Systems (ITSC) 1035–1040. https://doi.org/10.1109/ITSC.2018.8569650.
Roche, F., Brandenburg, S., 2020. Should the urgency of visual-tactile takeover requests match the criticality of takeover situations? IEEE Trans. Intell. Veh. 5 (2), 306–313. https://doi.org/10.1109/TIV.2019.2955906.
Roche, F., Somieski, A., Brandenburg, S., 2018. Behavioral changes to repeated takeovers in highly automated driving: effects of the takeover request-design and the non-driving related task-modality. Hum. Factors J. Hum. Factors Ergon. Soc. 61 (5), 839–849. https://doi.org/10.1177/0018720818814963.
Roche, F., Thüring, M., Trukenbrod, A.K., 2020. What happens when drivers of a highly-automated vehicle take over control in critical brake situations? Accid. Anal. Prev. 144 https://doi.org/10.1016/j.aap.2020.105588.
Rodemerk, C., Habenicht, S., Weitzel, A., Winner, H., Schmitt, T., 2012. Development of a general criticality criterion for the risk estimation of driving situations and its application to a maneuver-based lane change assistance system. 2012 IEEE Intelligent Vehicles Symposium 264–269. https://doi.org/10.1109/IVS.2012.6232129.
Rohrmann, B., 1978. Empirische Studien Zur Entwicklung Von Antwortskalen für die sozialwissenschaftliche Forschung. Zeitschrift Für Sozialpsychologie 9, 222–245.
SAE International, 2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016). https://www.sae.org/standards/content/j3016_201806/.
Salthouse, T.A., 1991. Mediation of adult age differences in cognition by reductions in working memory and speed of processing. Psychol. Sci. 2 (3), 179–183. https://doi.org/10.1111/j.1467-9280.1991.tb00127.x.
Sepehr, M., 1988. NASA Task Load Index: Deutsche Version (No. 284; pp. 1–16). Technische Universität Berlin. https://www.tib.eu/de/suchen/id/TIBKAT%3A019690789/NASA-Task-Load-Index-deutsche-Version/#.
Shahid, A., Wilkinson, K., Marcu, S., Shapiro, C.M., 2011. Karolinska sleepiness scale (KSS). In: Shahid, A., Wilkinson, K., Marcu, S., Shapiro, C.M. (Eds.), STOP, THAT and One Hundred Other Sleep Scales (Pp. 209–210). Springer New York. https://doi.org/10.1007/978-1-4419-9893-4_47.
Siebert, F.W., Oehl, M., Pfister, H.-R., 2014. The influence of time headway on subjective driver states in adaptive cruise control. Transp. Res. Part F Traffic Psychol. Behav. 25, 65–73. https://doi.org/10.1016/j.trf.2014.05.005.
van Nes, N., Brandenburg, S., Twisk, D., 2010. Improving homogeneity by dynamic speed limit systems. Accid. Anal. Prev. 42 (3), 944–952. https://doi.org/10.1016/j.aap.2009.05.002.

Vogel, K., 2003. A comparison of headway and time to collision as safety indicators. Accid. Anal. Prev. 35 (3), 427–433. https://doi.org/10.1016/S0001-4575(02)00022-2.
Wainer, H., Braun, H.I., 2013. Test validity. Routledge. https://books.google.de/.
Zhang, B., de Winter, J., Varotto, S., Happee, R., Martens, M., 2019. Determinants of take-over time from automated driving: a meta-analysis of 129 studies. Transp. Res. Part F Traffic Psychol. Behav. 64, 285–307. https://doi.org/10.1016/j.trf.2019.04.020.

D Publication 2: What happens when drivers of highly automated vehicles take over control in critical brake situations?

Accident Analysis and Prevention 144 (2020) 105588

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

What happens when drivers of automated vehicles take over control in critical brake situations?

Fabienne Roche*, Manfred Thüring, Anna K. Trukenbrod

Technische Universität Berlin, Germany

ARTICLE INFO

Keywords:
Driver behavior
Human-automation interaction
Driving simulator
Vehicle automation
Critical driving situation

ABSTRACT

Even with automated vehicles, driving situations with short time headways and extreme vehicle dynamics may arise when unpredictable events occur. If drivers take back control under such conditions, it is uncertain how they behave and how well they can cope with the situation. This issue has not been investigated yet and is subject to our study. In a driving simulator, non-distracted participants ($N = 42$) experienced nine critical situations caused by a braking vehicle in front of them. Time headway and longitudinal vehicle dynamics were varied to create different degrees of objective criticality. Participants' criticality ratings, take-over behavior, and driving performance were recorded and analyzed. The results indicate that participants were sensitive to changes in objective criticality and adapted their behavior. Take-over times were very fast under all conditions and participants showed higher criticality ratings, more intense decelerations, and more lane changes with increasing objective criticality. To avoid a collision, participants decelerated much more than the automation and changed lanes, even though this was not necessary. Thereby, they raised the risk of vehicle instability, rear-end collisions, and collisions with overtaking vehicles. To conclude, take-overs in critical brake situations may be a threat to the safety of drivers and other road users because drivers' reactions are more pronounced than necessary. These results suggest that assistive functions are required to support drivers in critical take-over situations.

1. Introduction

After almost a decade of intense psychological and ergonomic research on automated vehicles, it is time to ask what important issues have not been sufficiently addressed yet. One of these issues concerns the consequences of transitions of control under safety-critical conditions which result from short time headways and high vehicle dynamics. More precisely, in this paper, we investigate what might happen if drivers decided to take over control under such critical conditions in a brake situation.

An overview of the state-of-the-art reveals that an impressive amount of knowledge has already been accumulated about human factors in automated vehicles. For instance, experiments in driving simulators showed that people are very likely to take up *non-driving related tasks* (NDRTs) when the automation is switched on (Carsten et al., 2012; Naujoks et al., 2016). In line with these results, other studies showed that people are less alert during automated driving (Carsten et al., 2012; Saxby et al., 2013) and that NDRTs divert attention away from the traffic situation (Large et al., 2017). This decrease of situation awareness may become a hazard when drivers have to regain control on short notice. Take-over times and performance after the transition from automated to manual driving have been investigated concerning many factors, such as different NDRTs (Merat et al., 2012; Roche et al., 2018; Zeeb et al., 2017), different degrees of traffic density (Gold et al., 2016; Radlmayr et al., 2014), and different types of take-over requests (Politis et al., 2014; Roche et al., 2018). Eriksson and Stanton (2017) reviewed 25 experiments on takeover times mostly in urgent situations and reported a range from 1.14 s to 15 s.

To sum up, the studies show how the cognitive state of the driver, the surrounding traffic, and the design of the take-over request affect driving behavior and experience. Hence, psychological and ergonomic research has already provided much information about the transition of control from the automated vehicle to the driver. The leading experimental paradigm of this research is to investigate transitions which (1) must be handled by a *distracted* driver, (2) are *initiated* by the automated vehicle via a take-over request (system-initiated), and (3) take place under *uncritical conditions* concerning timing and vehicle dynamics.

Despite the valuable insights that were obtained with this paradigm,

* Corresponding author at: Technische Universität Berlin, Secr. MAR 3-2, Chair of Cognitive Psychology and Cognitive Ergonomics, Marchstraße 23, 10587, Berlin, Germany.

E-mail address: fabienne.roche@freenet.de (F. Roche).

https://doi.org/10.1016/j.aap.2020.105588
Received 20 January 2020; Received in revised form 8 April 2020; Accepted 7 May 2020

one issue has been fairly neglected until now: the potential consequences of transitions of control which are *initiated by the driver* instead of the automation. This kind of transition is called "driver-initiated"[1] as opposed to "system-initiated" take-over (McCall et al., 2016). According to the Amendment of Article 8 of the Vienna Convention on Road Traffic, such take-overs must be possible at all times (United Nations Economic Commission for Europe, 2014). However, surprisingly little is known about them, since only a few research studies have addressed this topic.

One of these studies was conducted as a field trial with 38 participants who drove a Tesla Model S equipped with an automation for longitudinal and lateral control (Schott et al., 2018). They were allowed to enable and disable the automation at any time while traveling on a circular course near Berlin, Germany, which included sections of a highway and rural roads. It was observed that participants frequently disabled the automation without being prompted by a take-over request.

Another study was conducted in a driving simulator with 36 participants (Epple and Brandenburg, 2018). The reliability of the system was varied by exposing the participants to different numbers of automation failures (0, 1 or 2) in the pre-phase of the experiment. In the experimental phase, the participants were instructed to regain control whenever they felt necessary. The phase started in the automated mode and the participant's vehicle sped up to 100 km/h. At this speed, it followed a lead vehicle keeping a distance of 0.50 s, 0.75 s, 1.00 s or 1.25 s time headway. The number of driver-initiated take-overs was surprisingly high over all conditions and amounted to 85% in total. Due to this ceiling effect, neither reliability and criticality led to significant differences in the number of take-overs, nor did reliability affect self-reported trust. It must be noted, however, that driver-initiated take-overs even took place when time headway was short, though they were less frequent in these cases.

As these two studies show, drivers take over control rather frequently and they may do so even under critical conditions, like short time-headways. This may become important in a number of safety-critical situations. To illustrate this point, imagine the following scenario: While you are traveling in the automated mode, a vehicle passes your car with a high velocity, squeezes into the lane between you and a leading vehicle and brakes heavily. The automation of your vehicle reacts immediately and brakes too to avoid a collision – but is the braking strong enough? Do you trust the automation to handle this unusual, hazardous situation? If not, you are likely to take over control by stepping on the brake pedal thus deactivating the automation. Even if you were not distracted by a NDRT in this situation, you might fail to manage the situation better than the automation or you might even cause a rear-end collision because you do not apply a sufficient braking force.

In a driving simulator study, we investigate such brake situations by deliberately deviating from the leading experimental paradigm with respect to three conditions: (1) Drivers are *not distracted* during traveling in the automated mode, (2) drivers must *react fast* if they want to regain control of the vehicle, and (3) drivers have to cope with take-over conditions of *varying objective criticality*. Different degrees of criticality are obtained by two parameters: Time headway to a vehicle in front and the vehicle dynamics of the participant's vehicle.

To our knowledge, this approach is unique and has not been pursued before. The central issue that is thus addressed is: *How do drivers perform when they take over control under highly dynamic conditions of varying criticality?* No other study has investigated this question yet. However, a review of seven manufacturers' tests of autonomous vehicles on public roads in California impressively demonstrated the great importance of 'autopilot disengagements' (Lv et al., 2018). The authors distinguish passive disengagements from active disengagement. While the former ones are initiated by the automation, the latter ones are initiated by the driver. As the review shows, the reasons for active disengagements are manifold. Some of them result from unexpected or reckless behavior of other drivers, or come close to our braking scenario, such as 'The driver felt a delay in deceleration, so the driver took over the brake operation.' (Lv et al.,2018, p. 61). Others are quite different, such as 'The automated vehicle moved uncomfortably close to a parked car.' (p. 61). As Lv et al. (2018) conclude: 'Disengagement with driver take-over is a complex process which is crucial to driving safety and ride comfort during automated driving.' (p. 65). From this perspective, our study provides new knowledge about the transition of control from the automation to the driver which may inspire the development of assistive systems that aim to increase the safety and comfort of such transitions.

1.1. Two parameters of objective criticality: time headway and vehicle dynamics

Time headway (THW) is defined as the time until the ego vehicle reaches the current position of a reference object (Vogel, 2013). A reference object may be a preceding vehicle or a stationary hazard, such as an obstacle on the road. THW depends on the distance to the reference object $d_{reference}$ and the velocity of the ego vehicle v_{ego} (see Eq. 1). The shorter THW is, the more critical is the situation.

$$THW = \frac{d_{reference}}{v_{ego}} \qquad (1)$$

Vehicle dynamics are affected by lateral and longitudinal forces (F_x and F_y). These forces add up to a horizontal force and influence the traction between road and tire (numerator in Eq. 2). As long as the actual horizontal force is smaller than the maximal possible force, the vehicle is stable. The maximal possible force depends on the friction coefficient μ_f and the vertical forces F_z acting on the tire (denominator in Eq. 2). The ratio of the actual and the maximal possible force describes the stability of the vehicle on the road and is termed traction usage (TU; Rajamani, 2011; see Eq. 2).

$$TU = \frac{\sqrt{F_x^2 + F_y^2}}{\mu_f * F_z} \qquad (2)$$

The maximal possible value of TU is 1. A higher TU caused by increased lateral or longitudinal deceleration or acceleration can result in slippage (Rajamani, 2011) and may thus lead to a high risk of losing control over the vehicle. In critical brake situations, TU raises when the advanced emergency braking system (AEBS) applies a braking force to avoid an impending collision (SAE International, 2018; United Nations Economic Commission for Europe, 2018). In case drivers take over control from an automated vehicle at this moment, they are confronted with high decelerations. If they increase the braking force or start steering, the actual traction and thus TU increases as well. As a consequence, the vehicle may become unstable or even uncontrollable.

Combinations of different values of THW and TU constitute brake situations of varying objective criticality. Although take-overs in such situations may be hazardous, they must be possible at all times as demanded by the Amendment of the Vienna Convention on Road Traffic (United Nations Economic Commission for Europe, 2014). As long as legislation is not altered in that respect, several research questions arise from this circumstance.

1.2. Research questions

In the present study, we investigate how drivers behave and perform when they take over control in brake situations of different objective criticality established by varying values of THW and TU. Three research questions (RQ) are addressed in this context:

[1] We use the term 'driver-initiated' instead of 'user-initiated' as proposed by McCall et al. (2016).

RQ 1: Does the objective criticality impact the subjective criticality indicating that drivers can discriminate between the different degrees of criticality?

RQ 2: Does the objective criticality influence take-over behavior?

RQ 3: Do drivers deliver an appropriate performance when they take over or do they increase the risk by their reaction?

2. Method

To answer the research questions, we used an experimental setup corresponding to the scenario described in the introduction. Participants were traveling in the automated mode when a vehicle passed them at a high velocity and merged into the lane between their vehicle and a lead vehicle. As soon as the merging was completed, the vehicle in front started breaking and the automation of the participant's vehicle reacted by braking as well. We asked what would happen if drivers interrupted the automated maneuver by braking themselves, especially under varying conditions of criticality. To answer this question, we established specific values of THW and TU at the time of takeover. For this purpose, an acoustic cue was employed. It was presented to trigger a braking response by the participant for the intended combinations of THW and TU. Since we used a trigger, it must be emphasized that we did *not* elicit genuinely driver-initiated transitions of control, but investigated *what might happen if* drivers intervened under controlled conditions of varying criticality.

2.1. Participants

Forty-two persons (19 women, 23 men) between 21 and 66 years of age ($M = 29.93$ years, $SD = 9$ years) participated in the driving simulator study. All participants had to be German native or near-native speakers to follow the German instructions. It was required that they had been holding a driving license for a minimum of four years ($M = 11.38$ years, $SD = 7.23$, $Max = 35$ years). Twenty (48%) participants reported to have experience with advanced driver-assistance systems, such as adaptive cruise control or lane change assistance. Each participant received 10 €/hour or course credits as gratification. The experiment was approved by the ethics committee of the Department of Psychology and Ergonomics of Technische Universität Berlin, Germany, and its conditions comply with the tenets of the Declaration of Helsinki (1964). Participants gave their informed consent before the experiment started.

2.2. Driving simulator

The study was conducted in a fixed-based driving simulator at the Department of Automotive Engineering at Technische Universität Berlin (see Fig. 1). The vehicle was an Audi A4 equipped with a force-feedback steering wheel, active pedals, original control interfaces, and a motion seat. The tilt of the seat simulated centrifugal forces, deceleration, and acceleration. The implemented automation had an adaptive cruise control system and a lane-centering mimicking an automated vehicle at SAE-level 3 (SAE International, 2018). Three projectors and a curved canvas enabled a front viewing angle of 180°. One rear projector in combination with a second canvas was used to ensure the visibility of the traffic behind the vehicle via the rear-view mirror and the left side mirror. Driving scenarios were implemented with SILAB 5.0 from the Würzburg Institute of Traffic Sciences GmbH (SILAB, 2014). The experimental road was a two-lane highway. The model of the vehicle dynamics was calculated by the commercial program IPG CarMaker 6.0. It estimated TU based on the longitudinal and lateral decelerations and accelerations, gravity, and friction coefficient (see Eq. 2). Gravity and friction coefficient were held constant in this study. The decelerations and accelerations were perceptible by the pace of the optical flow in the simulation and by the tilt of the motion seat. Driving noise was presented via a loudspeaker and driving vibrations were transmitted by a subwoofer. Additional microphones and speakers were used to enable the communication between the experimenter and the participant.

2.3. Procedure

The experiment consisted of a familiarization phase lasting three minutes, a training phase with 10 trials and an experimental phase with 15 trials. In the familiarization phase, participants had the opportunity to get accustomed to the simulator by driving on a rural road with some curves and little traffic. During the training, they experienced braking in the manual and the automated mode in five trials each. In the manual trials, they accelerated the vehicle up to 130 km/h and then drove through a passage bordered by traffic cones. Upon a trigger, participants had to reduce their speed to 50 km/h as fast as possible and then stop at the end of the passage. A tone with a base frequency of 1200 Hz and a duration of 0.6 s served as the trigger. For the automated trials, participants were told that the automation would take over longitudinal and lateral control during the drive. The provided information conformed with the description of an automated vehicle which matches SAE-level 3 (SAE International, 2018). Participants accelerated the vehicle manually, but at 30 km/h they received the auditory message "Automation takes over now". Simultaneously, the automation was switched on and the participants had to take their hands off the steering wheel and their feet off the pedals. The vehicle continued accelerating to 130 km/h and drove through the passage. Upon the trigger, participants were instructed to take back control by braking. Again, they had to reduce their speed to 50 km/h and stopped at the end of the passage.

Prior to the experimental phase, participants were instructed to drive as safely as possible and to avoid accidents by all means. They were requested to take over vehicle control as fast as possible when they heard the trigger. It was emphasized that they should only resume control when the trigger was given, but not otherwise. This was supervised by the experimenter and called for if necessary. Each participant completed 15 trials, nine experimental ones (3 THW x 3 TU) and six filler trials. For each participant, both types were randomly combined to three groups, each consisting of three experimental and two filler trials to ensure an adequate ratio. The order within and between groups was randomized to prevent a carry-over effect. The experimental trials represented brake situations of varying objective criticality. Each drive started on the hard shoulder of a two-lane highway with a lead vehicle placed 50 m ahead on the right lane. Participants were instructed to accelerate manually and to follow the lead vehicle. The automation was switched on just like in the training trials. It continued to accelerate up to 180 km/h[2] and followed the lead vehicle with a time headway of 1.8 s. In each trial, there was a medium traffic density on the neighboring lane with 5–7 cars passing the participant's vehicle. Number, color, and model of these cars were varied over the trials. After 1:30 min, a vehicle passed by, merged into the lane between the participant's and the lead vehicle and braked (see Fig. 2). The distance between the participant's and the merging vehicle as well as the deceleration varied depending on the experimental conditions (see Experimental Design). At the moment the vehicle merged into the participant's lane, the AEBS of the participant's vehicle started braking and the trigger was presented. Upon the trigger, participants were instructed to take over control, finished the maneuver and finally stopped on the hard shoulder. In case they failed to respond to the trigger, the automation completed the maneuver instead without causing an accident. The experimental trials were complemented by six filler trials. No trigger was given in three of these trials when a vehicle merged into the participant's lane and three times a trigger was issued without a merging vehicle. Together, both kinds of fillers served to reduce the predictability of required responses. After each trial, the participants rated

[2] For the German sample, a velocity of 180 km/h is not unusual, because there is no speed limit on many sections of German highways.

Fig. 1. The driving simulator at the Department of Automotive Engineering at Technische Universität Berlin.

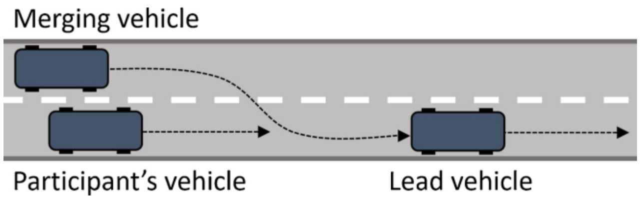

Fig. 2. Schematic illustration of the brake situation. The merging vehicle starts braking as soon as it has merged into the lane between the participant's and the lead vehicle.

the criticality of the experienced situation on a single item with a graphical analog scale, ranging from "not critical at all" (1 pt.) to "very critical"[3] (100 pt.). The item was presented on an iPad 1 with standard factory settings using the online survey service SoSci-Survey version 3.1 (www.soscisurvey.de).

At the end of the experiment, participants provided demographic information, such as age, gender and driving experience, on the iPad. All questions and questionnaires were presented in German. Overall, the experiment lasted about 1.5 h. After each trial, participants had the opportunity to take a break and were reminded of this option when they had completed eight trials.

2.4. Experimental design

Different degrees of objective criticality were achieved by two within-subject factors: time headway (THW) and traction usage (TU). THW was varied by the distance between the participant's and the merging vehicle and had three values: 0.34 s (least critical, 17.2 m distance), 0.21 s (medium critical, 10.3 m distance), and 0.08 s (most critical, 3.9 m distance). TU had three values as well: 0.44 (least critical), 0.64 (medium critical), and .84 (most critical). These values represent the maximal target TU-values which were reached by the AEBS. The target TU-values resp. the automation's deceleration were maintained for at least 2.5 s when the participants did not take over. Each participant experienced all nine brake situations which resulted from the combination of the THW- and TU-values.

The dependent variables consisted of the criticality rating, take-over time [s], maximal deceleration [m/s^2], lane change frequency [%], and collision frequency [%]. The criticality ratings served to investigate whether drivers could discriminate between different degrees of objective criticality (compare RQ 1). Take-over time was defined as the interval between trigger onset and driver reaction (steering or braking). Thresholds for a deliberate driver reaction were set to 10 N of brake pedal force and ± 1.4° of steering wheel angle, similar to Roche, Somieski, and Brandenburg (2018). To assess maximal deceleration, the highest value that was reached during braking was recorded per participant and trial. Take-over time and maximal deceleration indicated how drivers responded to the varying conditions of objective criticality (RQ 2). Short take-over times were regarded as indicators for a safe take-over behavior, overly strong decelerations as indicators for a hazardous take-over since they increased the risk of colliding with cars following in the rear. Trials were coded as collision when the participant's vehicle contacted the merging vehicle or the crash barrier. Lane changes and collisions pointed to situations that the driver could not handle adequately. They were used to judge the appropriateness of a take-over since lane changes were unnecessary and collisions could be avoided by adequate braking (RQ 3). In addition to lane changes and collisions, we computed the difference between the participant's maximal deceleration and the maximal deceleration that the automation would have reached without the participant's intervention. The automation was designed to decelerate strong enough to prevent a collision under all circumstances, but also to limit its deceleration to the required minimum. Hence, deviations from automated deceleration are suited to indicate suboptimal driver interventions (RQ 3).

2.5. Hypotheses and data analysis

We hypothesize that more critical brake situations lead to higher criticality ratings and more pronounced take-over behavior in terms of shorter take-over times and larger maximal decelerations, but also to

[3] Translation from German.

more lane changes and more collisions. The difference between the deceleration of the participants and the automation, as well as the interaction effects of THW and TU are analyzed exploratively.

Take-over time, maximal deceleration and deceleration difference are examined with three linear mixed effect models based on the "lme4"- package (Bates et al., 2015) in R version 3.4.1 (R Core Team, 2017). In addition, logistic models are fitted with the same package for the two binary variables lane change frequency and collision frequency. Degrees of freedom are estimated via Satterthwaite's method with the "lmerTest"-package (Kuznetsova et al., 2017). Linear mixed effect models are preferred over analyses of variances because they have a higher statistical power when missing values occur (Kliegl et al., 2011). The independent variables THW and TU are entered into the statistical models as linear predictors with a coding of -1 for the least critical value of THW resp. TU, a coding of 0 for the medium critical value, and a coding of 1 for the most critical value. Hence, the intercept of each model is an estimate of the grand mean. The slopes of the effects of THW resp. TU represent the estimated difference when these variables grow more critical by one value. In addition to the fixed effects, the model also includes a random intercept for each participant to account for interindividual differences and the repeated measure. The formula of the model is shown in Eq. 3.1. The distribution assumptions for the random intercept γj and the error term ε_{ij} are shown in Eqs. 3.2 and 3.3.

$$Y_{ij} = \alpha + \beta_1 * THW_i + \beta_2 * TU_i + \beta_3 * THW_i * TU_i + \gamma_j + \varepsilon_{ij} \quad (3.1)$$

$$\gamma_j \sim N(0, \sigma_\gamma^2) \quad (3.2)$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \quad (3.3)$$

Y_{ij} stands for the ith measurement value of person j for the dependent variable, α for the intercept (coded to represent the estimated grand mean). β_1 and β_2 represent the slopes for THW and TU, and β_3 the difference in slopes for the interaction of THW and TU. γj is the random intercept for each participant and ε_{ij} is the deviation of the measurement Y_{ij} from the prediction (error term that is not explained by the model).

For each model, its goodness-of-fit is represented by two coefficients: the marginal and the conditional coefficient of determination (R^2). The first one describes the proportion of variance explained by the fixed factors, the second one the proportion of variance explained by both, the fixed and random factors (Nakagawa and Schielzeth, 2013).

3. Results

Reactions are classified as missing when they occurred either before the trigger or 2.5 s after it. Thirteen single trials (3.4%) are excluded due to missing reactions and eight single trials (2.1%) due to measurement errors. After data selection, 357 (94.4%) experimental trials remain for analysis. The results of the linear mixed effects and logistic models are presented in Table 1. The main effects and significant interactions are shown in Fig. 3.

3.1. RQ 1: Subjective criticality

The criticality ratings were affected by both, THW and TU, similarly. With each increase of objective criticality due to either THW or TU, the ratings raised by approximately 8–9 rating points (see estimates for THW and TU on criticality ratings in Table 1 and Fig. 3a & 3b). No significant interaction between THW and TU was observed.

3.2. RQ 2: Take-over time and maximal deceleration

The intercept of take-over time indicates that participants took over very fast on average. Neither significant main effects of THW or TU on take-over time, nor a significant interaction were detected. This indicated that the experimental variations of THW and TU did not affect the speed of participants' take-over.

Significant main effects of both, THW and TU, on maximal deceleration[4] were observed, i.e., participants decelerated stronger, the more critical the situation was (see Table 1 for detailed results). Fig. 3c resp. 3d show the influence of THW resp. TU on maximal deceleration and the maximal deceleration of the automation (three dark grey horizontal lines in Fig. 3c and 3d). Additionally, a significant interaction effect of THW and TU was found. It indicates that the increase of deceleration due to increasing TU-criticality was more pronounced when THW was less critical (see Fig. 3e).

3.3. RQ 3: Deceleration difference, lane changes and collisions

To test whether participants decelerated unlike the automation, we analyzed the deceleration difference. The significant, positive intercept indicates that, on average, the participants' deceleration was stronger than the deceleration of the automation (see Table 1). The size of this deviation was not affected by THW, but strongly influenced by TU (see Fig. 3f and 3g). This effect was, however, part of an interaction with THW. When comparing the slopes of the THW-values, it is obvious that the decrease of the deviation due to more critical TU-values was strongest when THW was most critical (solid line in Fig. 3h).

Lane changes occurred in 52 of 357 trials (14.6%). Both, THW and TU had an effect on lane change probability.[5] This probability was higher, the more critical either THW or TU became (see Fig. 3i and 3j). The interaction of these variables had no significant effect. Since collisions occurred only in three out of 357 trials (0.8%, two with the crash barrier, one with the merging vehicle), no statistical analysis was performed. The two collisions with the crash barrier happened because participants steered strongly to the left, the collision with the merging vehicle because the participant reacted too late.

4. Discussion

As required by the Vienna Convention on Road Traffic, drivers must be able to take over control from the automation at all times (United Nations Economic Commission for Europe, 2014). They may do so for many reasons, e.g. when they are surprised by an unexpected event, when they misjudge the current vehicle dynamics or when they think that the automation may fail to manage a maneuver (compare Lv et al., 2018 for an overview of such reasons). In this study, we investigated how drivers behave and perform when they take over control in critical brake situations after a period of automated driving. More precisely, we examined whether the objective criticality impacted the subjective criticality (RQ 1), whether the objective criticality influenced participants' take-over behavior (RQ 2) and whether their take-over performance was appropriate (RQ 3).

4.1. RQ 1 and 2: Does the objective criticality influence subjective criticality and take-over behavior?

The results show that objective criticality influences subjective criticality (RQ 1) as well as take-over behavior (RQ 2). This conclusion is based on the effects of THW and TU on criticality ratings and maximal decelerations.

In line with our hypotheses, we found main effects of THW on criticality ratings and maximal decelerations. Higher objective criticality

[4] Trials resulting in a collision ($N = 3$ trials) were excluded from the analysis of deceleration, resulting in 354 trials included in the analysis. However, we also performed the analysis with all valid trials without observing any changes in the results.

[5] Logistic models calculate the probability of a certain event, e.g. a lane change, based on the observed frequency. Therefore, we use the term 'probability' instead of 'frequency' for reporting statistical results.

Table 1
Summary of statistics for time headway (THW) and traction usage (TU) per dependent variables. Note. The predictors THW and TU were coded with -1 for least, 0 for medium and 1 for most critical. Significance symbols: * p < .05, ** p < .01, *** p < .001.

Dependent variables	Statistics				
Criticality rating [1-100]	Estimate	Std. Error	df	t-value	p-value
Intercept (α)	49.21	2.65	40.81	18.58	< .001 ***
THW (β_1)	8.77	1.21	314.17	7.26	< .001 ***
TU (β_2)	8.16	1.21	314.20	6.76	< .001 ***
THW : TU (β_3)	−1.62	1.49	315.48	−1.09	.278
Variance explained: $R^2_{marginal}$ = 14.1 %, $R^2_{conditional}$ = 50.8 %					N_{trials} = 357
Take-over time [s]	Estimate	Std. Error	df	t-value	p-value
Intercept (α)	0.58	0.04	40.68	14.52	< .001 ***
THW (β_1)	−0.02	0.02	313.91	−1.62	.106
TU (β_2)	0.02	0.02	313.94	1.40	.162
THW x TU (β_3)	0.00	0.02	314.86	−0.06	.954
Variance explained: $R^2_{marginal}$ = 0.6 %, $R^2_{conditional}$ = 53.1 %					N_{trials} = 357
Maximal deceleration [m/s²]	Estimate	Std. Error	df	t-value	p-value
Intercept (α)	8.43	0.07	40.90	125.31	< .001 ***
THW (β_1)	0.29	0.05	312.22	5.37	< .001 ***
TU (β_2)	0.44	0.05	313.42	8.02	< .001 ***
THW x TU (β_3)	−0.19	0.07	315.79	−2.88	.004 **
Variance explained: $R^2_{marginal}$ = 20.2 %, $R^2_{conditional}$ = 31.0 %					N_{trials} = 354
Difference maximal deceleration [m/s²]	Estimate	Std. Error	df	t-value	p-value
Intercept (α)	1.80	0.07	41.01	26.70	< .001 ***
THW (β_1)	−0.06	0.06	312.45	−1.09	.277
TU (β_2)	−1.55	0.06	313.71	−27.37	< .001 ***
THW x TU (β_3)	−0.26	0.07	316.15	−3.66	< .001 ***
Variance explained: $R^2_{marginal}$ = 65.8 %, $R^2_{conditional}$ = 69.9 %					N_{trials} = 354
Lane change probability (logistic)	Estimate	Std. Error	z-value		p-value
Intercept (α)	−3.30	0.60	−5.54		< .001 ***
THW (β_1)	0.75	0.27	2.76		.006 **
TU (β_2)	1.07	0.28	3.79		< .001 ***
THW x TU (β_3)	0.49	0.33	1.50		.133
Variance explained: $R^2_{marginal}$ = 12.7 %, $R^2_{conditional}$ = 8.6 %					N_{trials} = 357

in terms of decreasing THW led to higher criticality ratings and higher maximal decelerations (see Figs. 3a and 3c). Based on the criticality ratings, we conclude that drivers can discriminate between different levels of objective criticality caused by THW. Shorter distances to the merging vehicle are more critical and should appear as more threatening. This is reflected in our participants' reactions. They responded with stronger decelerations when THW became more critical. Their behavior is an effective strategy to decrease the distance to the merging vehicle to a comfortable level. In order to reach this level, participants had to decelerate stronger when the distance was shorter, hence when THW was more critical.

In a way similar to THW, TU impacted criticality ratings and maximal decelerations. This is indicated by the observed main effects. When TU became more critical, criticality ratings increased and the maximal deceleration grew higher (see Figs. 3b and 3d). It seems that our participants were sensitive to the higher horizontal forces and/or the faster approaching of the merging vehicle due to its strong deceleration. Again, this is reflected in deceleration behavior. The participants' strategy to avoid a collision was to decelerate stronger when the merging vehicle decelerated stronger in the more critical TU-trials. In these trials, a strong response was indeed necessary to avoid a collision.

In addition to the reported main effects, an interaction of THW and TU occurred. When TU became more critical, the maximal deceleration increased more when THW was larger and, thus, less critical (see Fig. 3e). For the least critical THW-value (0.34 s), our participants adjusted their deceleration to the growth of TU (slope of the dotted curve in Fig. 3e). On the one hand, it seems that for more critical values of THW, their deceleration was less influenced by TU (slopes of dashed and solid lines in Fig. 3e). Hence, when TU was less critical, THW still made a difference, i.e. stronger decelerations occurred for shorter distances to the merging vehicle. On the other hand, when TU became more critical, a strong deceleration was necessary anyway and could barely get higher than 9 m/s^2 due to physical restrictions (Kudarauskas, 2007). Such a high level of deceleration typically constitutes an emergency braking. It was nearly reached under all conditions when TU was most critical and led to a ceiling effect, thus potentially concealing any influence of THW on deceleration that might have occurred.

Contrary to our hypothesis, no significant effects of THW and TU on take-over times were observed. This might have two reasons. First, it must be noted that the THW-values in our experiment were much smaller than in other studies (e.g., 3 s and 7 s in Roche and Brandenburg, 2018). Therefore, they might have been too minor to exert any influence. Second, the participants' take-overs were very fast with 567 ms on average, and even smaller than the minimal value reported in the overview by Eriksson and Stanton (2017). The rapidness of their response suggests that they closely followed the instruction to take over control as fast as possible when the trigger was presented. This resulted in a floor effect that may have obscured any influence of the two independent variables.

The effects we observed are partially in line with the results reported by Roche and Brandenburg (2018, 2019). The authors investigated the impact of THWs of 3 s and 7 s on take-over behavior. Similar to our current study, criticality ratings and brake input increased when time headways were shorter. These results on the relation between objective and subjective criticality conform with the outcome of a study by Hu et al. (2019) who investigated the influence of environmental factors on hazard perception. Hazard perception in traffic can be defined as the capability to anticipate dangerous situations on the road (McKenna et al., 2006). As Horswill and McKenna (2004) point out, it is closely related to a person's situation awareness (Endsley, 1995), specially to stage three, i.e., the prediction of upcoming events. In their experiment in a high-fidelity driving simulator, Hu et al. (2019) confronted 14 participants with 27 scenarios which differed with respect to three 'kinematic parameters` (p. 41): velocity and acceleration of the ego vehicle as well as its distance to an obstacle that occurred on the road. Hazard perception was assessed with a 7-point rating scale. It turned out that the interaction of the three factors impacted the hazard ratings. More precisely, only when the velocity was high (v = 100 km/

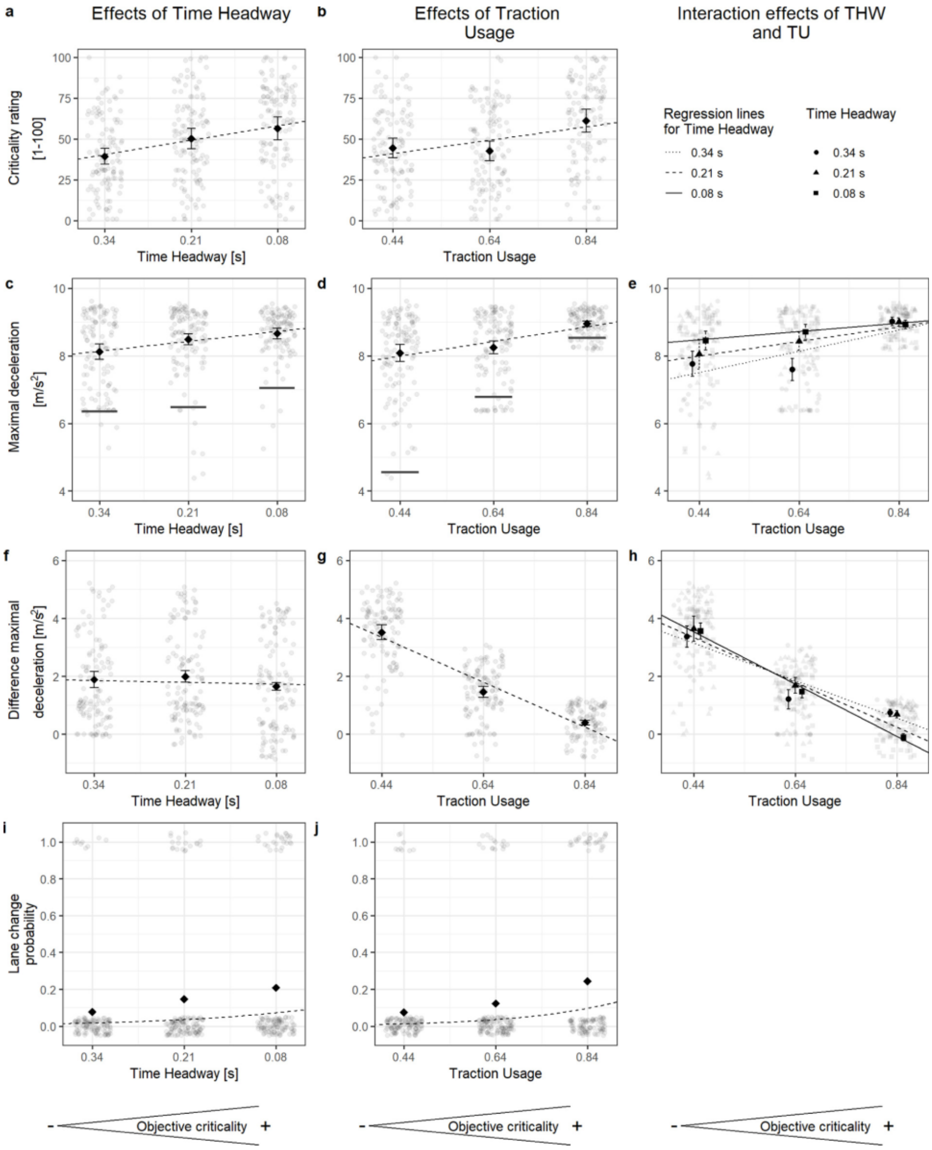

Fig. 3. Raw values, means per experimental condition, and regression lines for the dependent variables plotted for time headway (left), traction usage (middle), and their interaction (right). For each plot, the objective criticality increases from left to right. Only significant interactions are presented. The dark grey horizontal lines in the maximal deceleration plots (c and d) show the maximal deceleration by the automation. Lane change probabilities of 1 indicate a lane change. Note: Error bars stand for ± 1 standard error.

h), the interaction of acceleration and distance had a significant influence on the degree of the perceived hazard.

Obviously, the kinematic parameters in the study by Hu et al. (2019) share crucial features with the criticality factors of our experiment, especially with THW. Also, the results of the study point in the same direction as those of the other experiments. They demonstrate that drivers recognize changes of objective criticality and that critical parameters do not function independently from each other, but interact to constitute different degrees of experienced criticality and the specifics of hazard perception. Our current experiment expanded these insights by showing how these parameters additionally affect drivers' behavior when transitions of control occur during braking.

4.2. RQ 3: Do drivers deliver an appropriate performance when they take over or do they increase the risk by their reaction?

So far, the discussion has shown that (a) our participants were sensitive to changes of objective criticality, (b) reacted much faster than distracted drivers in other studies when a transition of control was demanded, and (c) decelerated with higher intensity when the situation was more critical. At first sight, it seems that their behavior was appropriate to the situation. However, a look at two other dependent variables reveals that their performance was partially suboptimal and riskier than necessary.

First, let us take a look at the disparity that shows when comparing our participants' deceleration with the deceleration of the automation[6]. The intercept of the deceleration difference represents the averaged difference between the participants' values and the values of the automation. Since the intercept is significantly higher than zero (see Table 1), it shows that our participants decelerated stronger than the automation (with about 1.8 m/s^2). Their behavior can be qualified as an overreaction because such intensity was not required to avoid a collision. Second, let us consider participants' lane changes. In 15% of the trials, lanes were changed, although, this was not needed to avoid a collision and poses the risk of rear-end collisions.

The interpretation of the deceleration difference and the lane change probability is complemented by the effects of THW and TU on these two dependent measures. For THW, a main effect on lane change probability was found (but not on deceleration difference). Participants adjusted their deceleration to more critical THW-values similar to the automation (almost horizontal slope in Fig. 3f). However, they decelerated 2 m/s^2 more on average than necessary (mean values in Fig. 3f). Additionally, more lane changes occurred when THW became more critical (see Fig. 3i). Since the distance to the merging vehicle was very short under more critical THW conditions, changing lanes probably appeared as an adequate strategy to avoid a collision when decelerating alone was perceived as not sufficient.

For TU, we observed two main effects. The deceleration difference decreased and lane change probability increased significantly when TU became more critical (see Fig. 3g and 3j). In brake situations with less critical TU-values, a weaker deceleration of 4.4 m/s^2 would have been sufficient, nevertheless, participants decelerated nearly as strongly as under the other conditions (~8 m/s^2, see Fig. 3d). Apparently, they considered this extent of deceleration as necessary to avoid a collision in the less critical TU-condition. However, the lane change probability increased when TU became more critical (see Fig. 3j). It seems that our participants choose a different behavior – strong braking alone or supplemented by lane changes – depending on the level of TU.

In addition to the main effects, an interaction of THW and TU affected the deceleration difference. When TU became more critical, it decreased stronger with more critical THW (slopes in Fig. 3h). In situations with the least critical TU (0.44), THW had a slight influence on the deceleration difference in the previously observed direction, i.e. more critical THW led to higher deceleration. But in take-over situations with the most critical TU, this effect was reversed because participants changed lanes more frequently, instead of decelerating when THW became more critical, too. Potentially due to that, weaker decelerations occurred in this situation.

The observed overreaction regarding deceleration difference is hazardous for two reasons. First, an overly strong deceleration increases TU and may destabilize the vehicle when the maximal possible force is approached. This can result in road departures or collisions with the crash barrier. Second, high decelerations may raise the risk of rear-end collisions with vehicles following behind. Therefore, a weaker, but more constant deceleration would be more appropriate.

Apart from the deceleration difference, the observed effects of the lane change probability are noteworthy because drivers may not check the neighboring lane carefully enough, thus, risking a collision with a passing car. Moreover, when steering occurs during braking, TU increases further and raises the probability of vehicle instability, loss of control, and accidents. The two collisions with the crash barrier can be traced back to strong steering in the last moment to avoid a collision with the braking vehicle.

To conclude, participants overreacted and showed worse take-over performance than the automation. Their deceleration was either inadequately strong or they even changed lanes without need. But how relevant are situations as investigated in our experiment at all for driving performance? At SAE-level 3, drivers are no longer obliged to monitor the situation. Many of them will use this new freedom to take up non-driving-related tasks, e.g., reading a newspaper or watching a video (Carsten et al., 2012). One might argue that such distracted drivers are never confronted with our experimental scenario, because they fail to recognize the hazard and the automation will handle the situation without any complication[7]. While this may be true in some cases, in others, the drivers' peripheral attention will enable them to notice a dangerous object outside their visual focus, especially when it is moving (Juola, 2016). However, distraction limits situation awareness (Young et al., 2013) and may delay responses since distracted drivers require additional time to update their mental model of the situation (Bellet et al., 2012). Such delays can provoke reactions that are even more pronounced than those we have observed in our experiment, because braking late requires a stronger braking force to prevent a collision. Therefore, the scenario that we investigated constitutes a situation that may become relevant for distracted drivers as well.

4.3. Limitations

The generalizability of the reported results may be restricted due to five experimental limitations:

1. Although our primary research interest concerns driver-initiated take-overs, the participants in our study did not decide themselves when to regain control over the vehicle. Instead, they responded to a trigger. Triggering was necessary to control different degrees of objective criticality that resulted from the combination of varying values of THW and TU. Triggered take-overs, however, may lead to different behavior and performance than driver-initiated take-overs.
2. Since our study included hazardous situations, it was conducted in a driving simulator. Take-over behavior in real traffic – particularly under critical and dynamic conditions – might differ from the one that is observed in a simulator. However, a comparative study by Eriksson et al. (2017) showed that a driving simulator is a valid tool to investigate the transition of control between driver and automation.
3. It must be emphasized that the driving simulator was static. Although a motion seat was employed to generate longitudinal and lateral forces, the success of this measure might be questionable. It is not clear how well these forces were perceived and to which degree they evoked realistic impressions. However, changes of velocity were as well reflected by changes of the optical flow, so that the visual channel also provided information about deceleration and acceleration. Nevertheless, future research should validate our results in a high-fidelity simulator or under real physical conditions in a protected space without traffic.
4. Nine brake situations were tested in a row. Compared to everyday traffic, this is a high frequency and may appear to be unrealistic. There is evidence that the repeated experience of similar situations – as in our study – leads to behavioral changes (Roche et al., 2018;

[6] The comparison also includes those values which the automation would have achieved if it had not been interrupted by the participants' take-over.

[7] We would like to thank an anonymous reviewer for this critical argument.

Zhang et al., 2019) and may bias experimental findings. In particular, it appears as likely that our participants got trained by the trigger and that the repetitions helped them to anticipate when and how to react. This might explain their ability to avoid collisions even under highly critical conditions.

5 The experiment was restricted to braking. Other critical situations resulting from THW or TU may require steering or a combination of steering and braking. More research is needed to assess the potential hazards of take-overs for these maneuvers.

5. Conclusion and future work

In contrast to most other studies on transition of control, our study did *not* investigate the performance of distracted drivers who respond to a take-over request when vehicle dynamics are uncritical. Instead, we focused on situations in which transitions of control were mastered by *undistracted* drivers who had to cope with *different degrees of criticality* resulting from time headway and traction usage during braking.

Since our participants were attentive and had the opportunity to train the take-over in the course of the experiment, we probably investigated the best case for braking under critical conditions during resp. after a take-over. This might be the reason for the low number of collisions. Nevertheless, our participants performed worse than the automation. On average, they decelerated more than necessary and some of them changed lanes without a cause. These overreactions are potentially dangerous since they might destabilize the car and increase the risk of collisions. Due to the limitations of our study, however, it remains open to what extent these results can be generalized and whether other maneuvers related to a take-over may also entail hazards.

Against this background, future research should address the following issues. First, studies should create circumstances which elicit *real* driver-initiated take-overs. This way, it would be investigated how often drivers interfere with the automation *on their own choice*, instead of responding to a trigger. While our study was restricted to the question what *might* happen if drivers take over control under critical conditions, such experiments would show whether and how often they actually do so and how well they manage the situation. However, this will be methodologically challenging because for the variation of criticality, neither THW nor TU can be strictly controlled without a trigger. In addition to mere braking, future studies should also investigate situations which require steering or combinations of braking and steering. Although ethical reasons prohibit to carry out such experiments in real traffic, it should be conducted under more realistic conditions, e.g. in a dynamic instead of a static driving simulator.

In our own research, we are currently planning a number of further studies. Some will take place in a dynamic driving simulator, others in a mixed reality environment. In this environment, participants will use an automated test vehicle on a wide, disused airfield. While driving in the automated mode, virtual objects on a collision course will be displayed. The automation will start to evade the obstacles, but participants will have to decide whether they trust their vehicle enough to handle the situation. As in our simulator study, THW and TU will constitute the objective criticality of the maneuvers.

The results of our studies will be used by automotive engineers at Technische Universität Berlin to develop an assistance system for driver-initiated take-overs. The basic idea for this system is to permit such take-overs, but to intervene when parameters like THW or TU approach critical values. Depending on the current behavior and maneuver, this could be accomplished by attenuating or intensifying the driver's input, e.g. by de- or increasing the braking force or by reducing or enlarging the steering angle. First concepts of assistance for evading maneuvers have been proposed by Nguyen and Müller (2019). However, more research is yet required to develop a comprehensive assistance system for driver-initiated take-overs.

CRediT authorship contribution statement

Fabienne Roche: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Visualization, Project administration. **Manfred Thüring:** Writing - review & editing, Project administration, Funding acquisition. **Anna K. Trukenbrod:** Writing - review & editing.

Declaration of Competing Interest

None.

Acknowledgements

This work has resulted from the interdisciplinary research project "Analysis and Support of Driver Interventions in Dynamic: Critical Situations during Highly Automated Driving" funded by the German Research Community (Deutsche Forschungsgemeinschaft, DFG, Grant No. 326727090). The study has been planned, enabled, and conducted in cooperation with Dr. Stefan Brandenburg, Sandra Epple, Jan Haentjes, Prof. Dr.-Ing. Steffen Mueller, Thang Nguyen, and Philipp Wittke.

References

Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48.

Bellet, T., Bornard, J.-C., Mayenobe, P., Paris, J.-C., Gruyer, D., Claverie, B., 2012. Computational simulation of visual distraction effects on car crivers' situation awareness. In: Rußwinkel, N., Drewitz, U., van Rijn, H. (Eds.), Proceedings of the 11th International Conference on Cognitive Modeling, pp. 59–65.

Carsten, O., Lai, F.C.H., Barnard, Y., Jamson, A.H., Merat, N., 2012. Control Task Substitution in Semiautomated Driving: Does It Matter What Aspects Are Automated? Hum. Factors J. Hum. Factors Ergon. Soc. 54 (5), 747–761. https://doi.org/10.1177/0018720812460246.

Endsley, M.R., 1995. Toward a theory of situation awareness in dynamic systems. Hum. Factors J. Hum. Factors Ergon. Soc. 37 (1), 32–64.

Epple, S., Brandenburg, S., 2018. Driver-initiated take-Over behavior in critical driving situations in highly automated driving. In: Human Factors and Ergonomics Society Europe Chapter Annual Meeting. Berlin.

Eriksson, A., Banks, V.A., Stanton, N.A., 2017. Transition to manual: comparing simulator with on-road control transitions. Accid. Anal. Prev. 102, 227–234. https://doi.org/10.1016/j.aap.2017.03.011.

Eriksson, A., Stanton, N.A., 2017. Takeover time in highly automated vehicles: noncritical transitions to and from manual control. Hum. Factors J. Hum. Factors Ergon. Soc. 59 (4), 689–705. https://doi.org/10.1177/0018720816685832.

Gold, C., Körber, M., Lechner, D., Bengler, K., 2016. Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density. Hum. Factors 58 (4), 642–652.

Horswill, M.S., McKenna, F.P., 2004. Drivers' hazard perception ability: situation awareness on the Road. In: Banbury, S., Tremblay, S. (Eds.), A Cognitive Approach to Situation Awareness: Theory and Application (Pp. 155–175). Ashgate Publishing.

Hu, Q., Guo, G., Wang, H., Tang, M., Lv, C., Cao, D., 2019. How to define the passenger's hazard perception level by combining subjective and objective measures? In: Stephanidis, C. (Ed.), HCI International 2019 – Late Breaking Papers (Vol. 11786, Pp. 40–53). Springer International Publishing. https://doi.org/10.1007/978-3-030-30033-3_4.

Juola, J.F., 2016. Theories of focal and peripheral attention. In: Bakker, S., Hausen, D., Selker, T. (Eds.), Peripheral Interaction (Pp. 39–61). Springer International Publishing. https://doi.org/10.1007/978-3-319-29523-7_3.

Kliegl, R., Wei, P., Dambacher, M., Yan, M., Zhou, X., 2011. Experimental effects and individual differences in linear mixed models: estimating the relationship between spatial, object, and attraction effects in visual attention. Front. Psychol. 1, 238. https://doi.org/10.3389/fpsyg.2010.00238.

Kudarauskas, N., 2007. Analysis of emergency braking of a vehicle. Transport 22 (3), 154–159. https://doi.org/10.3846/16484142.2007.9638118.

Kuznetsova, A., Bruun Brockhoff, P., Haub Bojesen Christensen, R., 2017. lmerTest: Tests in Linear Mixed Effects Models. https://CRAN.R-project.org/package=lmerTest.

Large, D., Banks, V., Burnett, G., Baverstock, S., Skrypchuk, L., 2017. Exploring the behaviour of distracted drivers during different levels of automation in driving. In: International Conference on Driver Distraction and Inattention. Paris, France.

Lv, C., Cao, D., Zhao, Y., Auger, D.J., Sullman, M., Wang, H., Dutka, L.M., Skrypchuk, L., Mouzakitis, A., 2018. Analysis of autopilot disengagements occurring during autonomous vehicle testing. Ieee/caa J. Autom. Sin. 5 (1), 58–68. https://doi.org/10.1109/JAS.2017.7510745.

McCall, R., McGee, F., Meschtscherjakov, A., Louveton, N., Engel, T., 2016. Towards a taxonomy of Autonomous vehicle handover situations. In: Automotive'UI. Proceedings of the 8th International Conference on Automotive User Interfaces and

Interactive Vehicular Applications 16. pp. 193–200. https://doi.org/10.1145/3003715.3005456.
McKenna, F.P., Horswill, M.S., Alexander, J.L., 2006. Does anticipation training affect drivers' risk taking? J. Exp. Psychol. Appl. 12 (1), 1–10. https://doi.org/10.1037/1076-898X.12.1.1.
Merat, N., Jamson, A.H., Lai, F.C.H., Carsten, O., 2012. Highly automated driving, secondary task performance, and driver state. Hum. Factors J. Hum. Factors Ergon. Soc. 54 (5), 762–771. https://doi.org/10.1177/0018720812442087.
Nakagawa, S., Schielzeth, H., 2013. A general and simple method for obtaining *R* 2 from generalized linear mixed-effects models. Methods Ecol. Evol. 4 (2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x.
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 Traffic Psychol. Behav. 38, 67–82. https://doi.org/10.1016/j.trf.2016.01.011.
Nguyen, T., Müller, S., 2019. Driver interventions in critical situations during automated driving. Advances in Dynamics of Vehicles on Roads and Tracks. IAVSD, Gothenburg, Sweden.
Politis, I., Brewster, S., Pollick, F., 2014. Evaluating multimodal driver displays under varying situational urgency. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 4067–4076. https://doi.org/10.1145/2556288.2556988.
R Core Team, 2017. *R: a Language and Environment for Statistical Computing* (Version 3.4.1) [Computer Software]. R Foundation for Statistical Computing. https://www.R-project.org/.
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. Hum. Factors Ergon. Soc. Annu. Meet. Proc. 58 (1), 2063–2067. https://doi.org/10.1177/1541931214581434.
Rajamani, R., 2011. Vehicle Dynamics and Control. Springer Science & Business Media.
Roche, F., Brandenburg, S., 2019. Should the urgency of visual-tactile takeover requests match the criticality of takeover situations? Ieee Trans. Intell. Veh. https://doi.org/10.1109/TIV.2019.2955906.
Roche, F., Brandenburg, S., 2018. Should the urgency of an auditory-tactile takeover request match the situational criticality? Intelligent Transportation Systems (ITSC) 1035–1040. https://doi.org/10.1109/ITSC.2018.8569650.
Roche, F., Somieski, A., Brandenburg, S., 2018. Behavioral changes to repeated takeovers in highly automated driving: effects of the takeover request-design and the non-driving related task-modality. Hum. Factors J. Hum. Factors Ergon. Soc. 61 (5), 839–849. https://doi.org/10.1177/0018720818814963.
SAE International, 2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016).
Saxby, D.J., Matthews, G., Warm, J.S., Hitchcock, E.M., Neubauer, C., 2013. Active and passive fatigue in simulated driving: discriminating styles of workload regulation and their safety impacts. J. Exp. Psychol. Appl. 19 (4), 287–300. https://doi.org/10.1037/a0034386.
Schott, R., Brandenburg, S., Thüring, M., Schröder, F., Telle, F., 2018. Who is driving whom? Highly automated driving transitions from manual to automation in reality and the impact of situation, personality and the driving experience. In: Human Factors and Ergonomics Society Europe Chapter Annual Meeting. Berlin.
SILAB (Version 5.0), 2014. Computer software]. WIVW GmbH. https://wivw.de/de/.
United Nations Economic Commission for Europe, 2014. Report of the Sixty-eighth Session of the Working Party on Road Traffic Safety.
United Nations Economic Commission for Europe. (2018). ECE/TRANS/WP.29/GRVA/2019/5.
Vogel, M., 2013. Temporal evaluation of aesthetics of user interfaces as one component of user experience. In: Smith, R.T., Wuensche, B.C. (Eds.), Proceedings of the Fourteenth Australasian User Interface Conference (AUIC2013) (Vol. 139, Pp. 131–132). Australian Computer Society.
Young, K.L., Salmon, P.M., Cornelissen, M., 2013. Missing links? The effects of distraction on driver situation awareness. Saf. Sci. 56, 36–43. https://doi.org/10.1016/j.ssci.2012.11.004.
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. https://doi.org/10.1016/j.trf.2017.07.008.
Zhang, B., de Winter, J., Varotto, S., Happee, R., Martens, M., 2019. Determinants of take-over time from automated driving: a meta-analysis of 129 studies. Transp. Res. Part F Traffic Psychol. Behav. 64, 285–307. https://doi.org/10.1016/j.trf.2019.04.020.

E Publication 3: What Happens when Drivers of Automated Vehicles Take Over Control in Critical Lane Change Situations?

Transportation Research Part F: Psychology and Behaviour 84 (2022) 407–422

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

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

What happens when drivers of automated vehicles take over control in critical lane change situations?

Fabienne Roche *, Sandra Becker, Manfred Thüring

Technische Universität Berlin, MAR 3-2, Chair of Cognitive Psychology and Cognitive Ergonomics, Marchstraße 23, 10587 Berlin, Germany

A R T I C L E I N F O

Keywords:
Vehicle automation
Simulator study
Take-over situations
Lane change
Criticality
Driver behavior
Evasion maneuver

A B S T R A C T

According to legislation, take-overs initiated by the driver must always be possible during automated driving. For example, when drivers mistrust the automation to handle a critical and hazardous lane change, they might intervene and take over control while the automation is performing the maneuver. In these situations, drivers may have little time to avoid an accident and can be exposed to high lateral forces. Due to lacking research, it is yet unknown if they recognize the criticality of the situation and how they behave and perform to manage it. In a driving simulator study, participants ($N = 60$) accomplished eight double lane changes to evade obstacles in their lane. Time-to-collision and traction usage were varied to establish different degrees of objective criticality. To manipulate these parameters as required, participants were triggered to take over control by an acoustic cue. This setting shows what *might happen* if drivers disable the automation and complete the maneuver themselves. The results of the experiment demonstrate that drivers rated objectively more critical driving situations as more critical and responded to the hazard very fast over all experimental conditions. However, their behavior was more extreme with respect to decelerating and steering than necessary. This impaired driving performance and increased the risk of lane departures and collisions. The results of the experiment can be used to develop an assistance system that supports driver-initiated take-overs.

1. Introduction

During the past few years, much psychological and ergonomic research has been conducted on automated driving. One important issue of this research concerned the transition of control. 'A transition in automated driving is defined as the process during which the human-automation system changes from one driving state to another…' (Lu, Happee, Cabrall, Kyriakidis, & de Winter, 2016, p. 1). So far, the main focus of most studies has been on transitions initiated by the automation, so-called 'system-initiated' take-overs in contrast to 'driver-initiated' ones (Lu et al., 2016; Martens, Schieben, & Merat, 2007; McCall, McGee, Meschtscherjakov, Louveton, & Engel, 2016). Only very few studies have been concerned with the latter.

Moreover, the majority of studies focused on take-over situations in which drivers had "sufficient time […] to respond appropriately to the driving situation at hand" (SAE International. (2018), 2018, p. 24). For example, a review from Eriksson and Stanton (2017) on 25 studies showed that the most frequently used lead time from a take-over request (TOR) to a critical event ranged from 3 s to 7 s. These values – particularly the higher ones – constitute rather controllable and uncritical conditions. However, it cannot be excluded that take-over situations with shorter times to a critical event occur, e.g. in case of driver-initiated take-overs.

* Corresponding author.
E-mail addresses: Fabienne.roche@freenet.de (F. Roche), epple@tu-berlin.de (S. Becker), manfred.thuering@tu-berlin.de (M. Thüring).

https://doi.org/10.1016/j.trf.2021.11.021
Received 12 February 2021; Received in revised form 30 October 2021; Accepted 24 November 2021
Available online 5 January 2022
1369-8478/

To summarize, research has gathered notable knowledge about drivers' behavior and performance after *system-initiated* take-over in *uncritical* situations, but little is known about the consequences of *driver-initiated* take-overs under *critical* conditions. But is this issue important at all? We claim that it is for two reasons. First of all, the Amendment of Article 8 of the Vienna Convention on Road Traffic requires that automated functions of a vehicle 'can be overridden or switched off by the driver' (United Nations Economic Commission for Europe, 2014, p. 9). Second, take-overs may happen in critical and dynamic driving situations for various reasons. Drivers may intervene at any time because they are startled by an unforeseen event or because they doubt that the automation can manage the current maneuver. If this happens, how well can drivers manage the situation, and how does its criticality affect their take-over behavior and performance?

Roche, Thüring, and Trukenbrod (2020) investigated transitions from automated to manual control in critical driving situations that required braking. Forty-two participants took part in the experiment. They were travelling in an automated vehicle when another car moved into their lane and started braking heavily. The automation responded immediately and braked to avoid a collision. In this situation, the transition of control took place under varying conditions of objective criticality established by different values of time headway and traction usage, varied by longitudinal deceleration. Although the drivers were fully alert (and not distracted as in most other studies), they worsened the situation by reacting too strong, i.e., they increased the risk of an accident by very strong decelerations and unnecessary lane changes.

In other driving situations, transitions of control may occur under conditions which require steering. Imagine the driver of an automated vehicle who is traveling on the highway while the automated mode is activated. Suddenly, the vehicle ahead of the driver changes lanes and reveals the view onto a close obstacle in the lane. Two cars have crashed and block the right side of the highway. The automation of the driver's vehicle reacts instantly by steering to evade the obstacle. The vehicle lurches to the left towards the crash barrier. Surprised by the maneuver, the driver intervenes and thus deactivates the automation. But is the steering strong enough or on the contrary too strong? Do drivers try to stay on course or are they struggling to change the direction? In any case, they must cope with demanding, short TTCs and with traction usages, which are more extreme than they are used to. If they fail to manage this situation, an accident can hardly be avoided.

Is the drivers' behavior appropriate for the criticality of the driving situation or do they react stronger than necessary like in the study by Roche et al. (2020)? If they do: in which situations, do they need support to avoid the dangerous consequences of their strong reactions? To answer these questions, we confronted participants with double lane changes in a driving simulator after short periods of automated driving. As in the study by Roche et al. (2020), different degrees of objective criticality were established by varying time-to-collision and traction usage, here realized by lateral acceleration. For this purpose, participants were triggered by an acoustic cue to take over vehicle control when the intended combination of both parameters was met like in Roche et al. (2020). This was accomplished at a certain point of time, either during the first or the second lane change. Since the take-over was triggered, it must be emphasized that we did not investigate *real* driver-initiated take-overs. Instead, we addressed the question what *might* happen, if drivers took over control under certain critical and dynamic conditions.

1.1. Double lane changes and parameters of objective criticality

A double lane change is often used to assess driver performance and the handling of a vehicle (ISO 3888-2, 2011). It is a demanding maneuver that is likely to occur in real traffic when obstacles appear, or slow vehicles are traveling in the driver's lane. An analysis of reasons for automation disengagements on public roads revealed that a lane change was one of many reasons why drivers did take back control (Lv et al., 2018). Two kinematic parameters affect the objective criticality of this situation (Hu et al., 2019): time-to-collision and traction usage.

Time-to-collision (TTC) denotes the interval from a certain point of time, e.g. a TOR, until a collision would occur with a reference object. A reference object can be a vehicle travelling ahead (Vogel, 2003) or a stationary obstacle in the lane (Bosnak & Skrjanc, 2017). TTC is calculated by the quotient of the distance to the reference object $d_{distance}$, the velocity of the vehicle v_{ego} and velocity of the reference object $v_{reference}$ (see Eq. (1)). The smaller TTC gets, the more critical the situation becomes.

$$TTC = \frac{d_{distance}}{v_{ego} - v_{reference}} \tag{1}$$

Various studies were conducted focusing on the available time for a take-over: the time budget operationalized by TTC or time headway (THW) at the moment of the take-over request. They showed that early take-over requests improved the driver's performance in terms of fewer collisions, fewer lane changes, better lane-keeping, higher time-to-collisions, and more adequate decelerations (Mok et al., 2015; Roche et al., 2020; Roche & Brandenburg, 2018; Zhang, de Winter, Varotto, Happee, & Martens, 2019). Moreover, subjective criticality and workload were lower and the experienced comfort was higher. However, the findings also indicated that drivers take their time to follow an early TOR, thus increasing take-over time (Roche et al., 2020; Roche & Brandenburg, 2018; Zhang et al., 2019).

Besides TTC, traction usage contributes to the objective criticality of a lane change. Due to the lateral movement during a lane change, lateral forces F_y act on the vehicle (Rajamani, 2011). Together with longitudinal forces F_x, they constitute the actual horizontal force; the numerator in Eq. (2). The vehicle is stable as long as the actual horizontal force does not exceed the maximal possible force, which is calculated by the friction coefficient μ_f and the vertical forces acting on the tire F_z. Exceeding the maximal possible force leads to instability of the vehicle and loss of vehicle control. This is the denominator in Eq. (2). The relation of the actual and the maximal possible force is called traction usage (TU) or adhesion utilization (Nguyen & Müller, 2020) and characterizes the stability of the vehicle (Rajamani, 2011; see Eq. (2)).

$$TU = \frac{\sqrt{F_x^2 + F_y^2}}{\mu_f * F_z} \quad (2)$$

The higher TU gets, the more critical the situation becomes because the vehicle approaches the maximal TU-value of 1 from which it gets instable, and the driver may lose control. The lateral forces during a lane change vary and depend on the trajectory of the vehicle. A steeper trajectory is characterized by higher lateral forces which in turn lead to higher TU. Until now, no study focused on the effects of traction usage on take-over behavior or performance.

1.2. Research questions

Automated driving systems are designed to change lanes safely even under critical conditions. However, in case of a driver-initiated take-over, such conditions may turn into a dangerous challenge since average drivers are not used to the high lateral forces that are acting during this maneuver. It is uncertain, how the drivers behave, whether their behavior is adapted to the objective criticality, and whether they are skilled enough to master such challenges.

As stated above, this study investigated *what might happen* if drivers took over control during a double lane change of varying degrees of objective criticality resulting from TTC and TU. In accordance with the study by Roche et al. (2020), three research questions were addressed:

(a) Does the objective criticality impact the *subjective criticality* indicating that drivers can discriminate between different degrees of criticality?
(b) Does criticality influence the *take-over behavior*?
(c) Do drivers deliver an *appropriate performance* when they take over during a lane change or do they increase the risk of an accident by their reaction?

2. Method

For implementing a double lane change, we used an experimental setting close to the scenario described in the introduction. Our study complied with the tenets of the Declaration of Helsinki (World Medical Association, 1964) and was approved by the ethics committee of the Department of Psychology and Ergonomics, Technische Universität Berlin, Germany.

2.1. Participants

Sixty persons (28 women, 32 men) between 20 and 62 years of age ($M = 29.1$ years, $SD = 9.1$ years) participated in the experiment. To ensure that they properly understood the German instructions, they were all German native or near-native speakers. Each of them had been holding a driver license for at least three years ($M = 11.3$ years, $SD = 8.6$, $Max = 44$ years). Thirty-two participants (53%) reported driving less than 5,000 km per year, nineteen participants (32%) between 5,000 and 10,000 km, four participants (7%) between 10,000 and 20,000 km, and five participants (8%) more than 20,000 km. Each participant received 10 €/hour or course credits as gratification for the experiment. All of them gave their informed consent before the experiment started.

2.2. Driving simulator

A fixed-based driving simulator of the Department of Automotive Engineering at Technische Universität Berlin served as technical platform for the experiment. The same simulator had been used in the previous study by Roche et al. (2020) to investigate critical brake situations. The simulator consisted of an Audi A4 equipped with a force-feedback steering wheel, active pedals, original control

Fig. 1. The driving simulator at the Department of Automotive Engineering at Technische Universität Berlin, Germany.

interfaces, and a motion seat. The motion seat was used to generate longitudinal and lateral forces which corresponded to the visual simulation of the driving. The implemented automation matched an automated driving system at SAE-level 3 (SAE International. (2018), 2018). It consisted of an adaptive cruise control and a lane-centering system. A comparison of different approaches to design automotive functions for vehicles had shown that a 5th degree polynomial is best suited to perform lane changes safely and comfortably (Sledge & Marshek, 1997). This approach was used to configure the automation in this study. This was achieved through steering; braking did not occur. The threshold for deactivating the automation was set to 18 N of brake pedal force and ±1 Nm steering wheel torque for less critical TU-conditions, resp. ±2.5 Nm steering wheel torque for more critical TU-conditions. The different thresholds accounted for the stronger steering of the automation in the more critical TU-trials.

A front view angle of 180° was enabled by three projectors and a curved canvas (see Fig. 1). The rear view was realized by one rear projector and a second canvas behind the vehicle. Participants could observe the traffic behind them via the rear-view mirror and the left side mirror. The visualization was implemented with SILAB 5.0 from the Würzburg Institute of Traffic Sciences GmbH (SILAB, 2014). The vehicle model was calculated by the commercial driving dynamics program IPG CarMaker 6.0. According to Eq. (2), TU was estimated based on the longitudinal and lateral accelerations, gravity, and friction coefficient. The friction coefficient was not varied in this study. For a more detailed description of the driving simulator, see Nguyen and Müller (2019a). The experience of lateral accelerations was conveyed by the dynamic visualization of the driven trajectory and by the lateral tilt of the motion seat. Driving noise was presented via speakers and driving vibrations were transmitted by a subwoofer. To enable the communication between the participant and the experimenter, the vehicle and the control room were equipped with microphones and speakers.

2.3. Procedure

The experiment consisted of three parts: a familiarization phase of three minutes, a training phase with six trials and an experimental phase with 14 trials. In the familiarization phase, participants drove on a rural road with some curves and little traffic to accustom themselves to the simulator. A passage bordered by traffic cones formed the route for a double lane change (or evasion maneuver) in the training phase. The training was split in two parts. First, participants exercised driving and changing lanes manually in two trials. They accelerated the vehicle up to 80 km/h and drove through the passage while keeping the speed constant. At the end of the passage, they braked to a stop. In the second half, participants were introduced to the automated driving system at SAE-level 3 (SAE International, 2018) and received their instructions. They were asked to take their hands off the steering wheel and their feet off the pedals while the automation managed the longitudinal and lateral control. They were informed that they had to take over control when they heard an acoustic signal. As the trial started, the vehicle accelerated up to 80 km/h and drove through the passage. When the acoustic cue sounded, participants took over control. The cue had a base frequency of 1200 Hz and a duration of 0.6 s. It was presented in two trials before the first obstacle and in two more trials before the second obstacle. After the take-over, participants drove through the rest of the passage manually and stopped at its end.

Before the experimental phase, participants were instructed to take over control only when they heard the trigger - but not otherwise - to avoid unwanted or over-hasty reactions. It was emphasized that the take-over should be performed as fast and safely as possible. Each participant completed eight experimental ones and six filler trials in randomized order. It was precluded that more than two filler trials were presented in a row and that a filler occurred as last trial. In the experimental trials, participants were traveling in the automated mode following a lead vehicle in front of them on a two-lane highway at a constant speed of 80 km/h and with a constant time-to-collision of 2.1 s to the car in front. One to two other vehicles were following the participant's vehicle in the left or the right lane. The color and the model of all cars changed from trial to trial. Participants performed no non-driving related task, because non-distracted drivers are more likely to initiate a take-over. After driving about one minute, the vehicle changed lanes and revealed the view onto two crashed vehicles in the right lane (see Fig. 2). The automation of the participant's vehicle reacted immediately by steering to the left to evade the obstacle. After passing this obstacle, the lead vehicle and the automated ego vehicle performed a second lane change, now back to the right, to evade road works ahead, thus completing the double lane change maneuver (see Fig. 2). The obstacles were 100 m apart. Each was 9 m long and 3.7 m wide. The trajectory of the maneuver comprised two symmetric lane changes. For the systematic variation of objective criticality in this situation, the driver's take-over was triggered at specific points in time when the automation was still in control. At these points, defined values of TTC and TU were obtained (for the specific variation of TTC and TU see Section 2.4). An acoustic cue, which occurred either during the first or the second lane change, served as trigger. The trigger was presented during the lane change either before the first or before the second obstacle. In case the participants failed to take over, the automation completed the double lane change without causing an accident. In addition to the experimental trials, participants experienced six filler trials to reduce the predictability of the situation. In these trials, less critical double lane changes were performed by the automation and no trigger was presented.

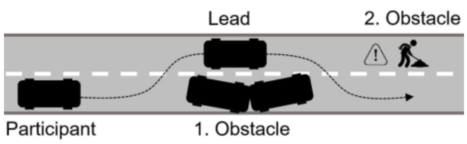

Fig. 2. Schematic illustration of the double lane change. *Note.* Two crashed vehicles constituted the first obstacle, construction works the second one.

At the end of each trial, the participant stopped the vehicle on the hard shoulder and then rated the experienced criticality of the situation by a single item (Roche, 2021). The item was implemented with SoSci-Survey version 3.1 (www.soscisurvey.de) and presented on an iPad 1. When all trials were completed, participants provided demographic information, such as age, gender and driving experience. The whole experiment lasted no more than 1.5 h, including breaks the participants requested.

2.4. Experimental design and hypotheses

TTC and TU served as within-subject factors in a 2 × 2 design. They were used to describe the criticality of the take-over situation. TTC and TU were experimentally varied by the timing of the trigger, i.e. the trigger was presented when the pre-defined values were reached. At this time, the automation was still active and, hence, TTC and TU were independent from the driver's behavior. The lane change was not treated as an independent variable because the two differed in many respects: the predictability of the obstacles (covered by the lead vehicle vs. early visible), the required driver response (steering to the left vs. steering to the right), and the event preceding the trigger (no lane change vs. first lane change). Therefore, data of the two changes were analyzed separately, based on the 2 × 2 design.

TTC was experimentally varied by the distance between the participant's vehicle and the obstacle at the moment of the trigger. For the first lane change, TTC was calculated regarding the first obstacle, for the second lane change, TTC was calculated regarding the second obstacle. TTC was varied on two levels: 2.1 s (less critical, 46.7 m) and 1.2 s (more critical, 26.7 m). The trigger was presented when the specific TTC-value was met and the automated driving system started the lane change. Therefore, the lane changes with the less critical (longer) TTC started earlier than the lane changes with the more critical (shorter) TTC (see Fig. 3).

TU was manipulated by the trajectory driven by the automation. When the automation of the participant's vehicle had to steer more, the trajectory was steeper, hence lateral acceleration and TU were higher (see Fig. 3). At the moment of the trigger, TU had one of two values: 0.24 (less critical) and 0.38 (more critical).

Each participant experienced all eight double lane changes, which resulted from the combination of the lane change (first or second), the TTC- (more or less critical), and the TU-values (more or less critical). For the whole lane change, it took the automation 2.25 s from the start of steering until driving straight again if the driver did not take over.

Dependent variables were the criticality rating, take-over time [s], maximal steering wheel angle [°], maximal (longitudinal) deceleration [m/s^2], lane departure frequency [%], and collision frequency [%]. The criticality ratings were assessed by a single-item rating scale ranging from 'not critical' (1pt.) to 'very critical'[1] (100pt.). Ten markings were inserted between the endpoints to provide orientation, similar to the NASA-TLX (Hart & Staveland, 1988). The scale had been developed and validated (Roche, 2021) and been used in previous studies (e.g., Roche et al., 2020). It served to answer *research question (a)* by capturing the effect of objective criticality (TTC and TU) on *subjective criticality*.

Take-over time was measured between the trigger onset and the driver's response by steering or braking. For maximal steering wheel angles and maximal deceleration, the highest value during the take-over was extracted per participant and trial. These maxima pointed out the extent of the participants' take-over behavior regarding lateral and longitudinal control. Large values indicated extreme and hazardous behavior. Strong steering increased the risk of destabilizing the vehicle and heavy decelerating the risk of causing rear-end collisions.

Take-over time, maximal steering wheel angles, and maximal deceleration served to investigate *research question (b)*. They represented the *take-over behavior* and indicated how participants reacted to the variations of TTC and TU.

A lane departure was registered when the vehicle crossed the right lane marking by at least 0.5 m. A collision of the participant's vehicle could occur with either the first obstacle, the second obstacle, the crash barrier, or the lead vehicle. Lane departures and collisions served as indicators of *take-over performance* in answer to *research question (c)*. Together with overly steering and strong deceleration, they pointed out situations that the driver could not handle adequately.

In line with the results of previous studies (Mok et al., 2015; Roche et al., 2020; Roche & Brandenburg, 2018; Zhang et al., 2019), we hypothesized that more critical conditions with respect to shorter TTCs and higher TUs would lead to higher criticality ratings and more extreme take-over behavior in terms of shorter take-over times, larger maximal steering wheel angles, and larger maximal decelerations. Regarding take-over performance, we expected more lane departures, more collisions as well as stronger steering and deceleration under the more critical conditions. Since we had no prior assumptions whether the two-way interactions between TTC and TU might affect the dependent variables, interaction effects were inspected exploratively.

2.5. Data analysis

For criticality ratings, take-over times, maximal steering wheel angles, and maximal decelerations, the main effects and two-way interactions were tested with linear mixed-effect models computed with the 'lme4'- package (Bates, Maechler, Bolker, & Walker, 2015) in R version 3.6.1 (R Core Team, 2019). We used linear mixed-effect models to analyze the results because they have a higher statistical power when missing values occur than analyses of variances (Kliegl, Wei, Dambacher, Yan, & Zhou, 2011). The assumption of the normal distribution of the residuals was tested with Shapiro-Wilk-test (Field, Miles, & Field, 2012). In case the residuals were not normally distributed, we transformed the dependent variable by Box-Cox-power-transformation with the 'MASS'-package in R

[1] Translated from German 'sehr kritisch'.

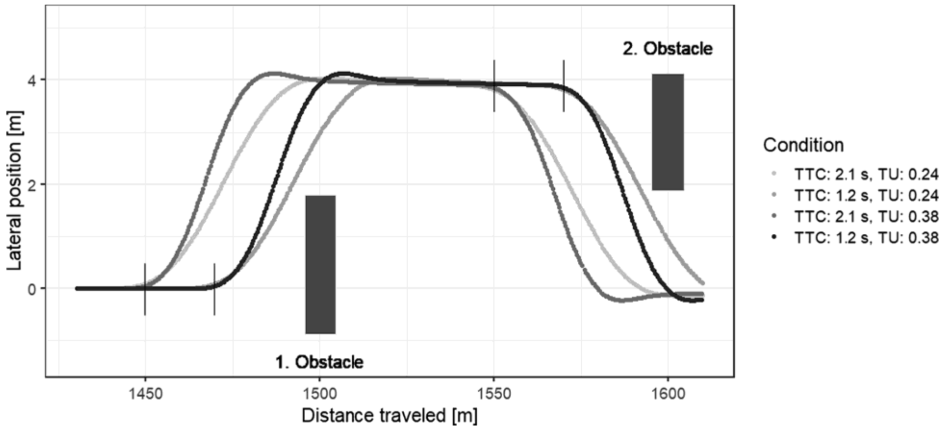

Fig. 3. Lateral position of the participant's automated vehicle. *Note.* If the participant had not taken over control, one of the four trajectories would have been driven by the automation, depending on TTC and TU. The vertical solid lines represent the trigger time during the first and the second lane change. The dark grey rectangles represent the locations of the obstacles.

(Venables & Ripley, 2002), log-transformation, square-root-transformation, and cube-root-transformation. Again, we tested with the Shapiro-Wilk-test whether the residuals of the transformed data were normally distributed. When transformed data were used to calculate the model, it is indicated in the corresponding results section and the Tables 3 and 6. These analyses were used to answer research question (a) concerning the effects of TTC and TU on criticality ratings as well as research question (b) concerning the effects of TTC and TU on take-over behavior. Logistic models were calculated with the same R-package for the binary variables, i.e. for lane departure and collision frequency. These analyses served to answer research question (c) concerning take-over performance. Degrees of freedom were estimated with Satterthwaite's method available in the 'lmerTest'-package (Kuznetsova, Bruun Brockhoff, & Christensen, 2017).

The independent variables TTC and TU were fixed variables and served as linear predictors with the coding of −0.5 for the less critical value and the coding of 0.5 for the more critical one. Due to the centering of the predictors, the intercepts α of the models represent the grand mean (see Eq. (4.1)). The estimates β_1 resp. β_2 constituted the estimated difference in the criteria when TTC resp. TU grew more critical from the less critical to the more critical value (research question a and b). The estimate β_3 represented the two-way interaction effects of TTC and TU. Besides these fixed effects, a random intercept for each participant γ_j was added to account for interindividual differences and the repeated measurement. Eqs. (4.2) and (4.3) show the distribution assumptions for the random intercept γ_j and the error term ε_{ij}.

$$Y_{ij} = \alpha + \beta_1 * THW_i + \beta_2 * TU_i + \beta_3 * THW_i * TU_i + \gamma_j + \varepsilon_{ij} \tag{4.1}$$

$$\varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right) \tag{4.2}$$

$$\gamma_j \sim N\left(0, \sigma_\gamma^2\right) \tag{4.3}$$

Note: Y_{ij} describes the i^{th} measurement of person j for the dependent variable. α reflects the intercept (grand mean), β_1 and β_2 the slopes for TTC and TU, β_3 the estimates for the interaction of TTC and TU. γ_j is the random intercept for each participant and ε_{ij} represents the error term that is not explained by the model.

In case a model failed to converge, the optimizer algorithm 'Bobyqa' (Powell, 2009) from the 'minqa'-package (Bates, Mullen, Nash, & Varadhan, 2014) was added to the model. The goodness-of-fit is represented by the marginal and conditional coefficient of determination (R^2). The marginal coefficient describes the variance explained by the fixed factors TTC and TU. The conditional coefficient characterizes the variance explained by the fixed and random factors (Nakagawa & Schielzeth, 2013).

In general, extreme steering or deceleration by the driver would cause high TU-values, thus, increasing the risk of losing control over the vehicle and of collisions with passing or following vehicles. These modes of behavior can be interpreted as a reaction stronger than necessary in the course of managing the situation after the take-over. Therefore, the evaluation of the take-over performance (research question c) investigated the appropriateness of steering and deceleration:

- The appropriateness of steering was examined by comparing the participants' maximal values to the calculated maxima that of the automation would have reached if no take-over had happened. Two-tailed, one sample t-tests were employed to detect significant differences between these values. The maximal steering wheel angle of the automation amounted to 27.1° for the less critical TU-trials and to 65.4° for the more critical ones[2]. The maximal values of the automation were higher in the more critical TU-trials due to the steeper trajectory in these trials.
- Since no deceleration was required to avoid a collision, the automation would have refrained from slowing down under any condition (see Section 2.2). Accordingly, two-tailed, one sample t-tests were calculated to determine whether the deceleration by the participants differed from 0.

Together, the t-tests on steering and deceleration revealed whether and in which situations participants reacted more extreme than necessary.

3. Results

The results chapter is divided into two sections, one for the first lane change (see Tables 1–3 and Fig. 4) and one for the second lane change (see Tables 4–6 and Fig. 5).

None of the residuals of the models were normally distributed, except those of the criticality ratings of the second lane change. For most models, the different transformation methods failed to approximate a normal distribution. In these cases, the non-transformed dependent variables were maintained, since model estimates based on non-transformed data are easier to interpret and robust to violations of distributional assumptions (Schielzeth et al., 2020). Only the take-over times in the first lane change and the maximal decelerations in the second lane change were transformed as indicated in the corresponding section.

As mentioned before, data for the first and the second lane change were analyzed separately. For both analyses, trials were excluded when take-overs had occurred either before the trigger or after the participants' vehicle had fully passed the obstacle (N = 24; 5%). And six experimental trials (1%) were excluded because of measurement errors. In total, 450 trials (94%) remained for analysis, 229 for the first and 221 for the second lane change.

3.1. First lane change

3.1.1. Subjective criticality

The analysis of criticality ratings revealed that higher TU led to significantly higher criticality ratings, while TTC did not affect the ratings (see Table 3). Also, the two-way interaction of both variables reached significance showing that the effect of TTC was more pronounced when TU was less critical (see Fig. 4a). The conditional variance explained by this model was far higher than the marginal variance, hence, a large portion of the variance was explained by the individual tendency of the participants to rate the criticality.

3.1.2. Take-over behavior

On average, participants took over after 0.82 s (SD = 0.26, min = 0.45, max = 2.22 s). For the statistical analysis, take-over times were transformed using the Box-Cox-power transformation ($\lambda = -1$). Neither TTC or TU nor their interaction had a significant effect on take-over times. In line with this result, the marginal coefficient of the model showed that TTC and TU hardly explained any variance, but rather the random intercept 'participant'.

TTC and TU significantly affected maximal steering wheel angles[3], such that the more critical TTC resp. TU led to higher steering wheel angles (see Fig. 4b and c).

The analysis of maximal deceleration revealed significant main effects of TTC and TU. With the more critical TTC resp. TU, participants' maximal deceleration was higher (see Fig. 4d and e).

3.1.3. Take-over performance

Lane departures occurred in 27 of 229 trials (11.7%, see Table 1 for detailed frequencies). Only TTC had a significant effect on lane departure probability.[4] Similar to the criticality ratings, the random intercept 'participant' added much to the explanatory value of the model (see the conditional coefficient in Table 3). This may be due to the small number of events (four trials) included in the model per participant.

In 30 of 229 trials (13.1%), participants collided with the first obstacle ($N = 26$) or the crash barrier ($N = 4$), but never with the second obstacle or the lead vehicle (see Table 2 for detailed frequencies of TTC and TU). None of the 26 collisions with the obstacle was a frontal one. Instead, our participants collided with the obstacle when they steered back to the right lane and had almost passed it, i.e., in real traffic they would have grazed the obstacle with their right back fender. The logistic model with the two-way interaction failed to converge, even when the optimizer was added. Therefore, a model without interaction was calculated. The probability of a collision

[2] The maximal steering wheel angles of the automation were derived from test trials without a driver take-over.

[3] Trials resulting in a lane departure ($N = 27$ trials) or a collision ($N = 30$ trials) were excluded from the analysis of maximal steering wheel angles and maximal deceleration. Including these trials, however, did not change any of the results.

[4] Logistic models estimate the probability of an event, e.g. a road departure, based on the observed frequency. Therefore, we use the term 'probability' instead of 'frequency' when reporting statistical results.

Table 1
Frequencies of lane departures in the first lane change related to TTC and TU.

Lane departure frequency	Less critical TTC	More critical TTC	Sum
Less critical TU	2	6	8
More critical TU	4	15	19
Sum	6	21	27

Table 2
Frequencies of collisions in the first lane change related to TTC and TU.

Collision frequency	Less critical TTC	More critical TTC	Sum
Less critical TU	2	0	2
More critical TU	26	2	28
Sum	28	2	30

Table 3
Summary of statistics for the first lane change: Main effects and interactions of time-to-collision (TTC) and traction usage (TU) on all dependent variables.

Criticality rating [1–100]	Estimate	Std. Error	df	t-Value	p-Value
Intercept	65.05	2.50	58.28	26.06	<0.001***
TTC	0.85	1.98	167.98	0.43	0.668
TU	22.46	1.99	168.63	11.30	<0.001***
TTC × TU	−10.54	3.98	168.69	−2.65	0.009**
Variance explained: $R^2_{marginal} = 19.9\%$, $R^2_{conditional} = 66.8\%$					$N_{trials} = 229$
Take-over time [s] (transformed)	Estimate	Std. Error	df	t-Value	p-Value
Intercept	−0.30	0.03	57.30	−9.57	<0.001***
TTC	−0.03	0.03	167.53	−1.03	0.305
TU	−0.06	0.03	168.48	−1.97	0.051
TTC × TU	−0.03	0.07	168.57	−0.40	0.693
Variance explained: $R^2_{marginal} = 1.3\%$, $R^2_{conditional} = 42.0\%$					$N_{trials} = 229$
Maximal steering wheel angle [°]	Estimate	Std. Error	df	t-Value	p-Value
Intercept	60.24	1.32	49.28	45.54	<0.001***
TTC	7.31	1.84	121.02	3.98	<0.001***
TU	35.47	1.85	126.07	19.13	<0.001***
TTC × TU	6.87	3.67	119.56	1.87	0.063
Variance explained: $R^2_{marginal} = 64.2\%$, $R^2_{conditional} = 74.2\%$					$N_{trials} = 177$
Maximal deceleration [m/s^2]	Estimate	Std. Error	df	t-Value	p-Value
Intercept	2.09	0.17	53.82	12.41	<0.001***
TTC	1.04	0.28	131.36	3.72	<0.001***
TU	1.60	0.28	137.30	5.71	<0.001***
TTC × TU	−0.09	0.56	129.82	−0.17	0.866
Variance explained: $R^2_{marginal} = 20.1\%$, $R^2_{conditional} = 31.7\%$					$N_{trials} = 177$
Lane departure probability (logistic)	Estimate	Std. Error		z-Value	p-Value
Intercept	−4.89	3.19		−1.53	0.126
TTC	2.57	1.27		2.02	0.043*
TU	1.62	0.92		1.77	0.077
TTC × TU	0.75	1.38		0.54	0.589
Variance explained: $R^2_{marginal} = 10.3\%$, $R^2_{conditional} = 57.0\%$					$N_{trials} = 229$
Collision probability (logistic)	Estimate	Std. Error		z-Value	p-Value
Intercept	−3.52	0.66		−5.31	<0.001***
TTC	−3.36	0.87		−3.85	<0.001***
TU	3.39	0.87		3.91	<0.001***
Variance explained: $R^2_{marginal} = 38.5\%$, $R^2_{conditional} = 41.1\%$					$N_{trials} = 229$

Note. The predictors TTC and TU were coded with −0.5 for the less and 0.5 for the more critical condition. Significance symbols: $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

was significantly higher when TU was more critical. Surprisingly, the collision probability was significantly higher when TTC was less critical.

The one sample t-tests showed that the participants' maximal steering wheel angles were significantly higher than those of the automation under three of four experimental conditions (df: 42–54, all $t > 7.0$, all $p < .001^{***}$). Only when TTC was less critical and

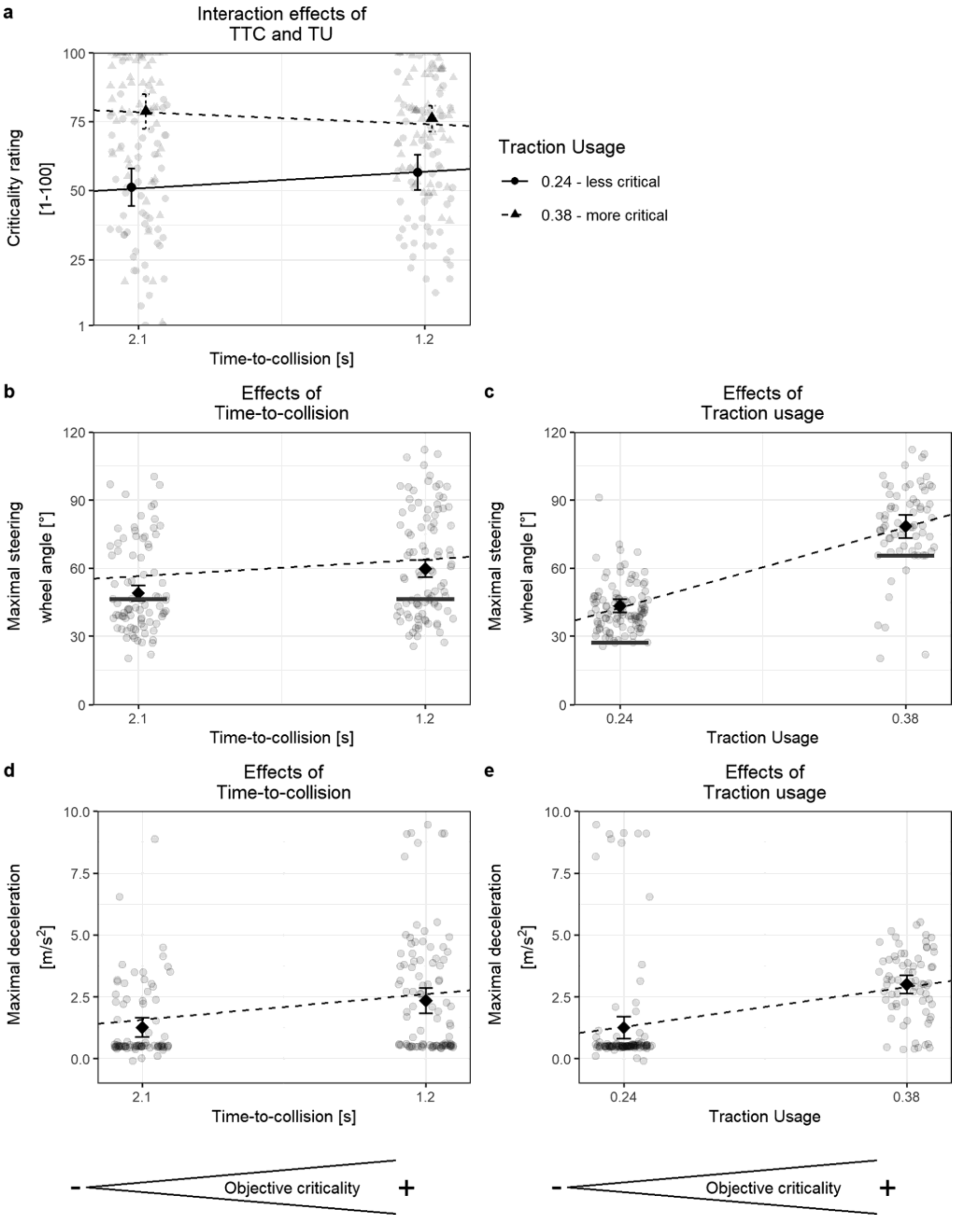

Fig. 4. Result plots of the dependent variables for time-to-collision (TTC), traction usage (TU), and their interactions for the first lane change. *Note.* Raw values, means per condition, and regression lines for each dependent variable are plotted. Only significant main effects and interactions are plotted and no main effects in case of a significant interaction. For each plot, the objective criticality increases from left t to right. The dark grey horizontal lines in the plots of maximal steering wheel angles (b and c) represent the maximal steering wheel angles the automation would have reached. Error bars represent ±1 standard error. A horizontal jitter with factor 0.1 was used to avoid an overlapping of the data.

Table 4
Frequencies of lane departures in the second lane change related to TTC and TU.

Lane departure frequency	Less critical TTC	More critical TTC	Sum
Less critical TU	1	20	21
More critical TU	4	5	9
Sum	5	25	30

Table 5
Frequencies of collisions in the second lane change related to TTC and TU.

Collision frequency	Less critical TTC	More critical TTC	Sum
Less critical TU	0	0	0
More critical TU	1	1	2
Sum	1	1	2

Table 6
Summary of statistics for the second lane change: Main effects and interactions of time-to-collision (TTC) and traction usage (TU) on all dependent variables.

Criticality rating [1–100]	Estimate	Std. Error	df	t-Value	p-Value
Intercept	56.45	2.52	58.31	22.36	<0.001***
TTC	17.11	2.14	162.40	8.00	<0.001***
TU	8.27	2.13	162.02	3.87	<0.001***
TTC × TU	−27.06	4.23	160.18	−6.39	<0.001***
Variance explained: $R^2_{marginal}$ = 20.3%, $R^2_{conditional}$ = 65.1%					N_{trials} = 221
Take-over time [s]	Estimate	Std. Error	df	t-Value	p-Value
Intercept	0.92	0.03	56.23	28.95	<0.001***
TTC	−0.16	0.03	160.34	−6.01	<0.001***
TU	0.02	0.03	159.95	0.57	0.566
TTC × TU	0.12	0.05	158.03	2.26	0.025*
Variance explained: $R^2_{marginal}$ = 8.3%, $R^2_{conditional}$ = 59.0%					N_{trials} = 221
Maximal steering wheel angle [°]	Estimate	Std. Error	df	t-Value	p-Value
Intercept	50.77	1.25	57.22	40.56	<0.001***
TTC	10.93	1.95	146.05	5.59	<0.001***
TU	29.27	1.95	143.37	15.02	<0.001***
TTC × TU	−10.00	3.88	139.99	−2.58	0.011*
Variance explained: $R^2_{marginal}$ = 56.4%, $R^2_{conditional}$ = 64.2%					N_{trials} = 190
Maximal deceleration [m/s²] (transformed)	Estimate	Std. Error	df	t-Value	p-Value
Intercept	−0.85	0.05	57.99	−18.78	<0.001***
TTC	−0.10	0.07	145.6	−1.44	0.152
TU	−0.34	0.07	142.99	−5.06	<0.001***
TTC × TU	0.17	0.14	139.75	1.26	0.208
Variance explained: $R^2_{marginal}$ = 7.2%, $R^2_{conditional}$ = 26.3%					N_{trials} = 190
Lane departure probability (logistic)	Estimate	Std. Error		z-Value	p-Value
Intercept	−2.63	0.46		−5.73	<0.001***
TTC	1.89	0.65		2.89	0.004**
TU	−0.09	0.65		−0.13	0.894
TTC × TU	−3.32	1.30		−2.56	0.010*
Variance explained: $R^2_{marginal}$ = 7.2%, $R^2_{conditional}$ = 26.3%					N_{trials} = 221

Note. The predictors TTC and TU were coded with −0.5 for the less and 0.5 for the more critical condition. Significance symbols: *$p < .05$, **$p < .01$, ***$p < .001$.

TU was more critical, their steering wheel angles did not differ from the automation. The steering wheel angles of the automation are represented by the horizontal lines in Fig. 4b and c.

Four one sample t-tests were calculated for maximal deceleration. They demonstrated that, participants decelerated under all conditions (df: 28–54, all $t > 3.5$, all $p < .001$***).

3.2. Second lane change

3.2.1. Subjective criticality

The mean rated criticality was 56.03 pts. (SD = 26.05, min = 1pt., max = 100 pts.). The model failed to converge, hence, the

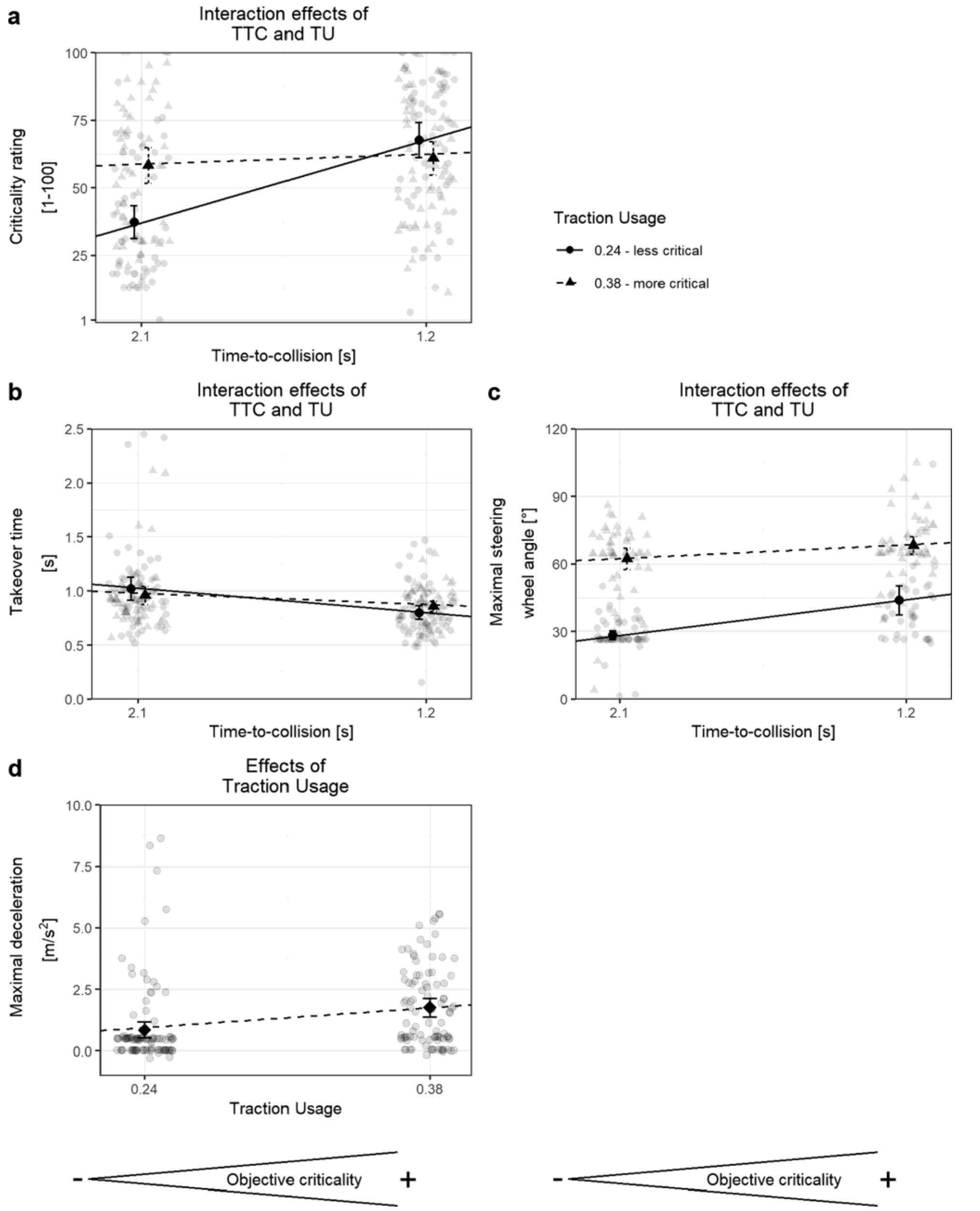

Fig. 5. Result plots of the dependent variables for traction usage (TU) and the interactions TTC × TU for the second lane change. *Note.* Raw values, means per condition, and regression lines for each dependent variable are plotted. Only significant main effects and interactions are plotted and no main effects in case of a significant interaction. For each plot, the objective criticality increases from left to right. Error bars represent ±1 standard error. A horizontal jitter with factor 0.1 was used to avoid the overlapping of data.

optimizer was added. The optimized model showed that TTC, TU, and their interaction had a significant effect on criticality ratings. The ratings increased when TTC resp. TU were more critical and the increase due to TTC was more pronounced with the less critical TU (see Fig. 5a). Again, the variance explained by TTC and TU was much smaller than the variance explained by those two fixed factors and the random intercept participant (see marginal and conditional coefficient in Table 6).

3.2.2. Take-over behavior

TTC and the interaction of TTC and TU had a significant effect on take-over time in such a way that the effect of TTC was stronger when TU was less critical (see Fig. 5b). However, not TTC or TU but the random intercept 'participant' contributed most to the explanatory value of the model (see marginal and conditional coefficient in Table 6).

The analysis of maximal steering wheel angles[5] revealed that they were significantly affected by TTC, TU, and their interaction. More critical TTC resp. TU led to higher maximal steering wheel angles. The effect of TTC was stronger when TU was less critical (see Fig. 5c).

The maximal decelerations were transformed using the cube-root-transformation. The maximal deceleration was only affected by TU. Participants decelerated more intensely when TU was more critical (see Fig. 5d).

3.2.3. Take-over performance

In 30 of 221 trials (13.6%), participants departed from the road to the right side (see Table 4 for detailed frequencies). The model showed that TTC and the interaction of TTC and TU had a significant effect on lane departure probability. More lane departures occurred when TTC was more critical.

In two of 221 trials (0.9%), participants collided with the second obstacle ($N = 1$) or with the crash barrier ($N = 1$, see Table 5 for detailed frequencies). These occurrences were too few for the calculation of a model.

Four one sample t-tests on maximal steering wheel angles demonstrated that the mean values differed from the steering behavior of the automation only under one condition ($t(39) = 5.245$, $p < .001$***). Participants' steering was significantly more extreme than that of the automation when TTC was more and TU less critical.

Although no deceleration was necessary to perform the second lane change, participants' average maximal deceleration was 1.34 m/s^2. Four one sample t-tests revealed that this behavior differed significantly from the automation under all conditions (df: 39–56, all $t > 3.5$, all $p < .001$***).

4. Discussion

Since drivers must have the option to regain control whenever they wish (United Nations Economic Commission for Europe, 2014), take-overs can happen in any situation. This study explored what might happen if driver-initiated take-overs occur during critical and dynamic double lane changes. We addressed three research questions, investigating (a) whether the objective criticality operationalized by TTC and TU impacts the *subjective criticality*, (b) whether the objective criticality affects the *take-over behavior*, and (c) whether participants show an appropriate *take-over performance*.

When discussing the results, one should keep in mind that the two parts of a lane change were different in various aspects, hence, a comparison is problematic. Besides, it should be noted that the distributional assumption was violated by most models which may limit the reliability of the results.

An interpretation of the observed interactions of the criticality ratings, take-over times, maximal steering wheel angles, and lane departures may be derived from a general characteristic of human attention, i.e., its limited capacity (Kahneman, 1973). When TU was critical, it constituted a hazard which was likely to capture the participants' attention. Since they were suddenly exposed to high changes of lateral forces and to extreme shifts of the optical flow in their field of view, they probably experienced the situation as very intense and menacing. Consequently, vestibular perception may have dominated visual perception and captured the participants attention. The processing of the experienced forces may have required more resources than usual, thus reducing the cognitive capacity for processing stimuli of other modalities, as the visual stimuli resulting from low TTC. On the other hand, when traction usage was less critical, more resources were available for appraising such information. In the first case, drivers may have been almost oblivious of any changes of TTC, while in the second case more capacity should have been available to recognize such changes and process them to judge the criticality of the situation. Hence, TTC could influence the perception of criticality and take-over behavior when TU was less critical, but not when it was more critical.

Of course, this is but a preliminary explanation. Together, the two criticality parameters constituted a complex pattern which calls for more studies to be fully understood. One branch of related research concerns human perception of hazards. For example, Hu et al. (2019) investigated the effects and interactions of the vehicle's velocity, acceleration (comparable with TU), and distance to an obstacle (comparable with TTC) on passengers' hazard perception. In consistence with our results, they found that the effects of these three parameters were not independent from each other but interacted and constituted the general impression of a hazard, especially when the velocity was high. However, they did not find a significant two-way interaction of acceleration and distance as we did.

An explanation for the weaker effects and interactions during the first lane change opposed to the second might be derived from the

[5] Trials resulting in a lane departure ($N = 30$ trials) or a collision ($N = 2$ trials) were excluded from the analysis of maximal steering wheel angles and maximal deceleration. Including these trials resulted in a highly significant interaction effect of TTC and TU on maximal steering wheel angles, and a significant effect of TTC and an interaction effect on maximal deceleration.

situational context. Due to the restricted visibility of the obstacle during the first lane change, the predictability of the course of events was constrained. Therefore, the situation as such might have appeared as so critical that a smaller TTC did not contribute much more to the experienced criticality. Or in other words, situational characteristics might have overruled the effect of TTC. This was not observed for the second lane change, where the obstacle was visible much earlier.

4.1. Research question (a): Does the objective criticality impact the subjective criticality indicating that drivers can discriminate between different degrees of criticality?

The observed significant effects showed that our participants consistently rated certain situations as more critical than others. There are two possible reasons for this finding. On the one hand, the participants might have truly evaluated the different criticalities of the driving situations at the moment of the trigger. In this case, they could have adapted their behavior to their perception of the situation, i.e. faster take-overs in situations that appeared as more critical. On the other hand, it could also be that the participants' success or failure in performing the maneuver might have influenced their judgements[6] since the ratings were completed *after* each trial. To check this assumption, we calculated correlations between criticality ratings and (a) take-over times ($r = -0.212$, $p < .001$***), (b) lane departures ($r = 0.087$, $p = .067$), and (c) collisions ($r = 0.245$, $p < .001$***). The first significant correlation showed that take-over times were lower when the subjective criticality was higher, indicating that drivers adapted their behavior to perceived criticality. This would support the assumption that participants really evaluated the criticality of the driving situation. The other two correlations revealed no significant relation between lane departures and ratings but between collisions and ratings. We can conclude that the a-posteriori assessment reflects the criticality of the driving situations but may also be biased by the outcome of the maneuver to some extent – at least when there was a drastic consequence such as a collision involved.

The ability to discriminate between less and more critical situations is crucial for acting adequately when taking over control under critical conditions. This leads to our second research question.

4.2. Research question (b): Does the objective criticality influence take-over behavior?

The reported main effects on take-over times, maximal steering wheel angles, and maximal decelerations indicated that the participants' take-over behavior was more extreme when TTC resp. TU were more critical. The effects could be explained by the general finding that more intense and aversive stimuli lead to more intense responses (Davis, 1984), i.e., in our experiment, critical values of TTC and TU caused more pronounced behavior in terms of the aforementioned dependent variables. Additionally, the shorter time that was available to act under the more critical TTC-conditions, may have provoked a steering stronger than necessary. This was not mandatory to avoid a collision (see the two horizontal lines in Fig. 4b), because trajectories of the maneuver under the less and the more critical TTC-condition were the same, and so were the required steering angles (see Fig. 3). Similarly, stronger decelerations were observed in the more critical TU-trials, although a higher deceleration was not required to manage these situations.

Since some of the participants' behavioral parameters were stronger than necessary, the question arises whether this implies an impairment of driving performance. This issue is addressed in the next section.

4.3. Research question (c): Do drivers deliver an appropriate performance when they take over during a lane change or do they increase the risk of an accident?

The effects on collision and lane departures frequencies, together with the different steering wheel angles and decelerations applied by the automation and the drivers showed that drivers tended to react stronger than the automation when they took over control from the automation (design of the automation see Section 2.2). They steered more extremely than the automation under most conditions during the first lane change and under one condition during the second lane change. Additionally, they decelerated when taking over under all conditions, even though no deceleration was necessary to perform the lane change. Strong steering and strong deceleration can lead to collisions with overtaking vehicles, the crash barrier, the obstacle, or rear-end collisions. It can also cause lane departures or vehicle instability when drivers steer or brake too much and thus further increase TU. Indeed, the mentioned potential consequences of such behavior, i.e. collisions and lane departures, occurred.

The total of 32 collisions and 57 lane departures is alarming since *none* would have occurred if the automation had completed the maneuver instead of the driver. In the 26 collisions with the first obstacle, participants strived to get back to the right lane as soon as they got aware of the second obstacle at the expanse of checking their surroundings. Indeed, the take-over times in these trials were shorter than in the other trials ($M_{frontal\ collision} = 784$ ms, $SD_{frontal\ collision} = 142$ vs. $M_{remaining\ trials} = 830$ ms, $SD_{remaining\ trials} = 273$ ms). Turning their head slightly to the right would have sufficed to recognize that it was too early to steer back.

Surprisingly, collisions were less likely when TTC was more critical in the first lane change. At first sight, this appears as paradox since usually a short TTC increases the risk of collisions. In this situation however, things were different. To establish the more critical value, the trigger was presented rather late. Therefore, most of the maneuver was carried out by the automation and the vehicle had passed most of the obstacle when the participants took over. This left hardly any chance to collide with the obstacle even when the participants' responses were fast and pronounced.

[6] We are grateful to an anonymous reviewer for making us aware of this possibility.

The effect of TTC on lane departures may be explained by the higher steering wheel angles that were also observed: Higher steering wheel angles increase the probability of lane departures.

4.4. Limitations

The study had three limitations that should be considered when interpreting our results:

1. Although our research interest concerned driver-initiated instead of system-initiated take-overs, our participants *did not decide themselves* to regain control. Instead, a trigger was used to elicit the transition of control from automation to the driver. This was necessary to create comparable lane change conditions with different degrees of objective criticality resulting from TTC and TU. Hence, we did not investigate what *happens* but what *might happen,* when drivers disengage the automation.
2. Since our study concerned with critical lane changes, it would have been irresponsible to conduct it under real traffic conditions. Behavior in a driving simulator, however, might deviate from behavior in the field. In our experiment, two conditions might have caused such deviations:
   - First of all, the driving simulator was static. Even though we used a motion seat to generate longitudinal and lateral forces, it is unclear how well these artificially created forces evoked realistic impressions of motion. However, the car movements were also conveyed by the visual presentation on the screen. The driven trajectory and the speed of the vehicle determined the drivers' optical flow so that lateral and longitudinal movements as well as decelerations and accelerations were visually perceptible. This may have compensated unrealistic vestibular information. Moreover, only two participants suffered from simulator sickness which indicated an adequate consistency of the vestibular stimulation with the visual information.
   - Secondly, most of the collisions occurred during the first lane change when our participants steered back to the right lane and contacted the obstacle which they had almost passed. Since the screen of the simulator provided a view of more than 180°, our participants' peripheral perception or a side glance should have enabled them to recognize that it was too early to move back into the right lane. Hence, it seems that the collisions resulted from the drivers' neglect of important visual information. Another factor, however, may have also contributed to the accidents: Since the driver's car was fixed in the simulator and did not really move, our participants may have failed to develop an adequate representation of its length. Therefore, it is not certain that drivers would behave the same way under real traffic conditions.
3. Eight similar take-over situations were tested in every session. This was a high frequency compared to realistic traffic conditions. Previous studies have shown that repeated exposure to take-over situations leads to changes in behavior and experience (Roche, Somieski, & Brandenburg, 2018; Zhang et al., 2019). Over the trials, our participants may have adapted their behavior to the take-over situation by learning when to expect them. Moreover, the repetition of the trigger may have trained them to respond immediately without spending time on decision making and response selection. Therefore, the small take-over times that we observed, may have resulted from our experimental conditions. They are not comparable to take-over times that are system-initiated or to those that occur in real traffic

All in all, the discussed limitations may have biased the results of our study in a certain direction. Since our participants were not distracted by non-driving related tasks, had the chance to form helpful expectations in the course of the experiment, and were trained by successive trials to perform the required response very fast, we may have investigated the *best case* of what might happen, if drivers took over control during double lane changes.

5. Conclusions and future work

The research questions and the experimental setting of our study differed from most others on transitions of control. To our knowledge, no one has yet varied TTC and TU to investigate drivers' subjective criticality, behavior and performance during take-overs except for Roche et al. (2020) in their driving simulator study on braking. So, what new insights have been gained by this approach and what follows from them?

The results on double lane changes showed that our participants were probably sensitive for higher objective criticality and appraised the situation accordingly[7]. They succeeded to take over control remarkably fast, but compared to the performance of the automation, the subsequent behavioral pattern carries features of a stronger reaction than necessary. This may be hazardous for the ego driver and passengers but also jeopardize other road users.

Comparable results were reported by Roche et al. (2020). The settings and the design of their experiment were the same as those of the present study, except for the investigated maneuver. Again, drivers were aware of different degrees of criticality, took over control very fast and reacted stronger than necessary. They braked heavier than necessary and changed lanes without cause. Collisions occurred too, but since the brake maneuver was less difficult than the double lane change, the number was much smaller (3 out of 357 trials; 0.8%).

Together the studies on steering and braking investigated what *might happen* if drivers took over control under dynamic and safety–critical conditions. This use case is important for traffic safety because drivers *must* always have the option to deactivate the

[7] As discussed in Section 4.1, the appraisal may have been biased to some extent by the participants' success or failure to complete the lane change flawlessly.

automation in order to regain vehicle control (United Nations Economic Commission for Europe. (2014), 2014). Even in dangerous situations, they might decide to initiate a take-over when it seems necessary. A variety of occasions can trigger such a decision. Drivers might be startled by the action of another road user, be surprised by an obstacle that suddenly occurs in their lane, or simply have the impression that the automation does not perform a maneuver appropriately. Whatever the occasion might be, the two studies demonstrated that a take-over under highly dynamic conditions entails overly pronounced behavior thus impairing the driving performance and increasing the risk of accidents (see also Roche et al., 2020). Since the limitations of both studies were very similar, impairments probably occurred for the *best case* of transitions from automated to manual driving. More severe consequences are to be expected when drivers do not get accustomed to critical take-overs by frequent exposures or when their situation awareness is reduced due to distraction.

Since it would be illegal to thwart driver-initiated take-overs completely as long as the Amendment of Article 8 of the Vienna Convention persists, safety can only be increased by an appropriate assistance system. The results of our studies have been used by automotive engineers at Technische Universität Berlin to develop the prototype of such a system. Their general approach is to intervene when TTC or TU reach critical values without suppressing driver-initiated take-overs. Depending on the maneuver, this is accomplished by adjusting the braking force or the steering angle to an appropriate value (Nguyen & Müller, 2019a, 2020). The results of the current study were employed to parameterize the steering assistance of the prototype.

More research, however, is yet necessary to convert the prototype into a comprehensive assistance system for driver-initiated take-overs. This development should be accompanied by further studies that evaluate the next versions of the system and expand the empirical basis for the assistance. They should investigate whether our findings can be replicated in a dynamic driving simulator or outside the laboratory under safe but physically realistic conditions. Especially experiments doing without a trigger are required to reveal under which conditions *real* driver-initiated take-overs occur.

CRediT authorship contribution statement

Fabienne Roche: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Sandra Becker:** Data curation, Writing – review & editing. **Manfred Thüring:** Funding acquisition, Supervision, Writing – review & editing.

Declaration of Competing Interest

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

Acknowledgements

This work results from an interdisciplinary research project ‘Analysis and Support of Driver Interventions in Dynamic, Critical Situations during Highly Automated Driving’. It was supported by the German Research Community (Deutsche Forschungsgemeinschaft DFG, grant number 326727090). The study has been planned, enabled, and conducted in cooperation with Dr. Stefan Brandenburg, B. Sc. Marc Buchholz, B. Sc. Jan Haentjes, Prof. Dr.-Ing. Steffen Müller, and M. Sc. Thang Nguyen.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.trf.2021.11.021.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Bates, D., Mullen, K. M., Nash, J. C., & Varadhan, R. (2014). *minqa: Derivative-free optimization algorithms by quadratic approximation* [R package version 1.2.4]. https://CRAN.R-project.org/package=minqa.

Bosnak, M., & Skrjanc, I. (2017). Efficient Time-To-Collision Estimation for a Braking Supervision System with LIDAR. In *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)* (pp. 1–6). https://doi.org/10.1109/CYBConf.2017.7985775

Davis, M. (1984). The Mammalian Startle Response. In R. C. Eaton (Ed.), *Neural Mechanisms of Startle Behavior* (pp. 287–351). Springer. https://doi.org/10.1007/978-1-4899-2286-1_10.

Eriksson, A., & Stanton, N. A. (2017). Takeover Time in Highly Automated Vehicles: Noncritical Transitions to and from Manual Control. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 59*(4), 689–705. https://doi.org/10.1177/0018720816685832

Field, A. P., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology, 52*, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hu, Q., Guo, G., Wang, H., Tang, M., Lv, C., & Cao, D. (2019). How to Define the Passenger’s Hazard Perception Level by Combining Subjective and Objective Measures?. In C. Stephanidis (Ed.), *HCI International 2019 – Late Breaking Papers* (Vol. 11786, pp. 40–53) Springer International Publishing. https://doi.org/10.1007/978-3-030-30033-3_4.

ISO 3888-2 (2011). *Passenger cars—Test track for a severe lane-change manoeuvre—Part 2: Obstacle avoidance*.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.398.5285&rep=rep1&type=pdf.

Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: Estimating the relationship between spatial, object, and attraction effects in visual attention. *Quantitative Psychology and Measurement, 1*, 238. https://doi.org/10.3389/fpsyg.2010.00238

Kuznetsova, A., Bruun Brockhoff, P., & Haub Bojesen Christensen, R. (2017). *lmerTest: Tests in Linear Mixed Effects Models*. https://CRAN.R-project.org/package=lmerTest.

Lu, Z., Happee, R., Cabrall, C., Kyriakidis, M., & de Winter, J. C. F. (2016). Human factors of transitions in automated driving: A general framework and literature survey. *Transportation Research Part F: Traffic Psychology and Behaviour, 43*, 183–198. https://doi.org/10.1016/j.trf.2016.10.007

Lv, C., Cao, D., Zhao, Y., Auger, D. J., Sullman, M., Wang, H., … Mouzakitis, A. (2018). Analysis of autopilot disengagements occurring during autonomous vehicle testing. *IEEE/CAA Journal of Automatica Sinica, 5*(1), 58–68. https://doi.org/10.1109/JAS.2017.7510745

Martens, Pauwelussen, Schieben, Flemisch, Merat, Jamson, & Caci (2007). *Human Factors' aspects in automated and semi-automatic transport systems: State of the art* (3.2.1; p. 55). European Commission.

McCall, R., McGee, F., Meschtscherjakov, A., Louveton, N., & Engel, T. (2016). Towards A Taxonomy of Autonomous Vehicle Handover Situations. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications – Automotive'UI 16*. https://doi.org/10.1145/3003715.3005456

Mok, B.-K.-J., Johns, M., Lee, K. J., Miller, D., Sirkin, D., Ive, P., & Ju, W. (2015). Emergency, Automation Off: Unstructured Transition Timing for Distracted Drivers of Automated Vehicles. In *Proceedings of the IEEE 18th Intelligent Transportation Systems (ITSC)* (pp. 2458–2464). https://doi.org/10.1109/ITSC.2015.396

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4* (2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nguyen, T., & Müller, S. (2019a). Driver Interventions in Critical Situations during Automated Driving. *Advances in Dynamics of Vehicles on Roads and Tracks*. Gothenburg, Sweden: IAVSD.

Nguyen, T., & Müller, S. (2020). Assistance Systems for Driver Interventions in Critical Situations during Automated Driving. *31th IEEE Intelligent Vehicles Symposium, Las Vegas, USA*.

Powell, M. J. D. (2009). *The BOBYQA algorithm for bound constrained optimization without derivatives (DAMTP 2009/NA06)*. Centre for Mathematical Sciences, University of Cambridge.

R Core Team. (2019). *R: A language and environment for statistical computing (3.6.1) [Computer software]*. R Foundation for Statistical Computing. https://www.R-project.org/.

Rajamani, R. (2011). *Vehicle Dynamics and Control*. Springer Science & Business Media.

Roche, F. (2021). Assessing subjective criticality of take-over situations after automated driving: Validation of two rating scales. *Accident Analysis & Prevention, 159*, 1–13. https://doi.org/10.1016/j.aap.2021.106216

Roche, F., & Brandenburg, S. (2018). Should the urgency of an auditory-tactile takeover request match the situational criticality?. In *Proceedings 2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1035–1040). https://doi.org/10.1109/ITSC.2018.8569650

Roche, F., Somieski, A., & Brandenburg, S. (2018). Behavioral Changes to Repeated Takeovers in Highly Automated Driving: Effects of the Takeover Request-Design and the Non-Driving Related Task-Modality. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 61*(5), 839–849. https://doi.org/10.1177/0018720818814963

Roche, F., Thüring, M., & Trukenbrod, A. K. (2020). What happens when drivers of a highly-automated vehicle take over control in critical brake situations? *Accident Analysis & Prevention, 144*. https://doi.org/10.1016/j.aap.2020.105588

SAE International (2018). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016)*. https://www.sae.org/standards/content/j3016_201806/.

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., … Sutherland, C. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution, 11*(9), 1141–1152. https://doi.org/10.1111/mee3.v11.910.1111/2041-210X.13434

SILAB (5.0) (2014). [Computer software]. *WIVW GmbH*. https://wivw.de/de/.

Sledge, N. H., Jr, & Marshek, K. M. (1997). Comparison of ideal vehicle lane-change trajectories. *SAE Transactions*, 2004–2027.

United Nations Economic Commission for Europe (2014). *Report of the sixty-eighth session of the Working Party on Road Traffic Safety*.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.

Vogel, K. (2003). A comparison of headway and time to collision as safety indicators. *Accident Analysis & Prevention, 35*(3), 427–433. https://doi.org/10.1016/S0001-4575(02)00022-2

World Medical Association (1964). *Declaration of Helsinki*. https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/.

Zhang, B., de Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation Research Part F: Traffic Psychology and Behaviour, 64*, 285–307. https://doi.org/10.1016/j.trf.2019.04.020

Further reading

Nguyen, T., & Müller, S. (2019b). Analyse und Unterstützung von Fahrereingriffen in dynamischen, kritischen Situationen beim hochautomatisierten Fahren. *VDI-Berichte, 2349*, 77–96.