Heliyon 10 (2024) e24112

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Getting back in the loop: Does autonomous driving duration affect driver's takeover performance?

Arthur Portron [a,*], Gaëtan Perrotte [a,b], Guillaume Ollier [a], Clément Bougard [b], Christophe Bourdin [a], Jean-Louis Vercher [a]

[a] *Aix Marseille University, CNRS, ISM, Marseille, France*
[b] *Groupe Stellantis, Centre Technique de Vélizy, Vélizy-Villacoublay, France*

ARTICLE INFO

Keywords:
Takeover request
Driver behavior
Long durations
Non-critical event
Simulating level-3 automation

ABSTRACT

The level 3 autonomous driving function allows the driver to perform non-driving-related tasks such as watching movies or reading while the system manages the driving task. However, when a difficult situation arises, the driver is requested to return to the loop of control. This switching from driver to passenger then back to driver may modify the driving paradigm, potentially causing an out-of-the-loop state. We tested the hypothesis of a linear (progressive) impact of various autonomous driving durations: the longer the level 3 autonomous function is used, the poorer the driver's takeover performance. Fifty-two participants were divided into 4 groups, each group being assigned a specific period of autonomous driving (5, 15, 45, or 60 min), followed by a takeover request with a time budget of 8.3 s. Takeover performance was assessed over two successive drives via reaction times and manual driving metrics (trajectories). The initial hypothesis (linearity) was not confirmed: there was a nonlinear relationship between autonomous driving duration and takeover performance, with one duration (15 min) appearing safer overall and mixed performance within groups. Repetition induced a major change in performance during the second drive, indicating rapid adaptation to the situation. The non-driving-related task appears critical in several respects (dynamics, content, driver interest) to proper use of level 3 automation. All this supports previous research prompting reservations about the prospect of car driving becoming like train travel.

1. Introduction

Technological advances over the last decade have reduced the gap between the familiar manually-driven car and the still unfulfilled promise of a fully autonomous, even flying, car. Particular attention is currently being paid to "autonomous" functions that can increase road safety and improve travel comfort. These new functions theoretically span a range of autonomy levels, from 0 for a fully manual car to 5 for a driverless, fully autonomous car [1]. So far, level 3, "*conditional driving automation*" is likely the most functional and deployable function on the market, given the current technological patents by manufacturers such as Mercedes-Benz or Honda [2, 3] and earlier ones by Audi, Volvo or Tesla for level 2+ driving functions [4–6].

Under the right conditions, this level of automation involves activating a system able to fully manage both vehicle speed and

* Corresponding author.
E-mail addresses: arthur.portron@gmail.com (A. Portron), gaetan.perrotte@stellantis.com (G. Perrotte), guillaume.ollier@cea.fr (G. Ollier), clement.bougard@stellantis.com (C. Bougard), christophe.bourdin@univ-amu.fr (C. Bourdin), jean-louis.vercher@univ-amu.fr (J.-L. Vercher).

https://doi.org/10.1016/j.heliyon.2024.e24112
Received 19 November 2023; Received in revised form 24 December 2023; Accepted 3 January 2024
Available online 22 January 2024

trajectory according to the road environment. During this phase, the driver should be able to spend his/her time on a wide range of non-driving-related tasks (NDRT, i.e., working, watching movies, playing games, texting). Nevertheless, when necessary (e.g., planned situation or unexpected conditions), the driver must be able to regain control within a safe transition time, in order to manage any event. Autonomous systems are designed to display a takeover request (TOR) demanding the driver's full focus on the driving task when the system assesses a need for it. A semi-autonomous vehicle thus requires drivers to develop new habits and routines, since conditional or highly automated levels change the range and even the nature of the interactions between the vehicle and its driver [7].

This involves a complex shift in role, from the classic "manual driver's" total involvement throughout the driving task to a role alternating between driver and passenger, or to an intermediate role that consists in "monitoring the proper functioning of the automated vehicle" [8]. Therefore, the system needs to assist the driver with the interaction and inform him/her about what the system is doing. This calls for research on human behavior, primarily to understand driver behavior under both manual and automated conditions, including the ability to extract information and any potential modulations of cognitive processes involved in semi-autonomous driving [9].

1.1. Out-of-the-loop states

Automation level 3 is commonly defined as "hands-off, eyes-off, mind on". This means that though the driver temporarily leaves the physical loop of vehicle control, he/she should must keep at least part of his/her cognitive resources focused to be attentive to a TOR displayed by the automatic driving system, and even on the environment in order to maintain a minimum level of his/her Situation Awareness (SA), and both in order to be able to regain control when requested (by TOR) or necessary (in case of possible false negative produced by the system). This condition is well documented in the field of human factors in Aeronautics [10], where, SA has been defined by Endsley [11] as " *the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future*". Applied to highly automated road vehicles, SA could be defined as the ability to stay in an intermediate on-the-loop state, to detect, and to correctly interpret a TOR signal, then quickly but safely get back into the loop of control of the vehicle.

Nevertheless, there are noticeable differences between aeronautic and automotive needs, for at least two reasons: 1) in terms of potential unexpected events, a road vehicle is generally moving within a more complex environment than a flying plane and 2) pilots are generally well trained to deal with autonomous systems under precise protocols, while by definition car drivers have so far not been trained to share control, and may react in various and even unpredictable ways when switching from one mode to the other. Consequently, unlike in Aeronautics, SA is difficult to sustain efficiently in autonomous car driving phases due to lack of specific driver expertise and training.

Moreover, in manual mode, by activating the steering wheel and pedals, the driver is involved in a physical loop of control of the vehicle, as well as in a cognitive loop of predicting, planning, and deciding what to do in the driving environment [12]. During level 3 or higher phases of automation, the driver switches voluntarily from the state of being "*in the loop*" to "*out of the loop*" (OOTL) through a potential intermediate "*on the loop*" state [13], whereas this latter framework is never involved in manual driving. It is now acknowledged that leaving the loops (of either cognitive or physical control or both) is critical [14]. It can negatively impact the driver's performance when collaborating with automated systems [15–17] and can result in a loss of SA, one of the deadliest forms of which is drowsiness [18]. Furthermore, in a monotonous driving environment, the risk of fatigue and loss of SA with automation is likely to be greater than under manual driving [19,20]. To facilitate collaboration between autonomous driving systems and human operators, it is crucial to identify factors inducing the "*out-of-the-loop*" state in an autonomous driving environment as well as those which facilitate getting "*back into the loop*".

1.2. Highly automated driving for long durations

Several factors influencing the transition back to manual driving after a takeover request have been extensively studied, particularly factors related to i) the human operator [21], like age [22,23], experience [24–26], level of fatigue [27,28]), ii) the driving environment, like weather [29], traffic [30], road type [28,31]), iii) non-driving-related tasks ([32–34], see reviews [35,36]). Studies have also explored parameters related to the takeover request itself (e.g., modality of TOR [37,38], time budget (i.e. the time between TOR and impending collision) [39,40], and duration of autonomous period before TOR [41,42]).

However, for practical reasons, most of these studies were designed to expose the driver to short periods of autonomous driving before a TOR, rarely involving their participants in really long autonomous driving phases. Feldhutter et al. (2017) [41] showed that a period of 20 min of automated driving increases response time to a takeover request but does not affect either acceleration or time-to-collision profiles. Bourrelly et al. (2019) [42] tested even longer periods (60 min vs 10 min) and showed that overall takeover performance was poorer after the longer period, with a longer reaction time and sharper trajectories during avoidance maneuvers. Vogelpohl et al. (2019) [20] furthermore designed a protocol to compare automated driving versus manual driving under a prolonged driving condition (60 min) or lack of sleep condition (<5 h sleep the night before the experiment). They confirmed that in a given driving environment, autonomous driving shortens the time before signs of fatigue appear. This effect was even more pronounced in some of the "lack of sleep" drivers, who took longer to deactivate the autonomous function than the "long drive" group. To summarize, though recent studies suggest that long autonomous driving phases could have deleterious effects on driving performance, there is no consensus on which driving parameters are impacted. It is therefore difficult to make strong recommendations to automotive manufacturers regarding whether automation should be extended to longer periods or, on the contrary, limited to shorter periods. The latter recommendation would run counter to the objectives behind highly automated functions.

1.3. Aim of the study

In order to characterize the dynamics of the influence of autonomous driving duration on driver's capabilities during takeover including an obstacle avoiding maneuver, and thus describe more precisely the effect of various durations on takeover performance, we conducted an experiment involving four independent groups. Each group was assigned a specific autonomous driving period (5, 15, 45, or 60 min). We questioned the existence of a potential progressive (linear) relationship between the duration of autonomous phases and takeover performance. If demonstrated, such a linear relationship would obviously make it easier to define a maximum (or optimal) duration of use of level 3. Also, since the participants would be exposed to an avoiding maneuver for the first time, each participant executed the protocol twice, to assess the influence of novelty (on the first trail) and afterward adaptation (during the second trial) on performance. We expected SA to be higher during the second trial, leading to better performance. The quality of takeover performance was assessed through several behavioral measures: reaction times, quality of control during the takeover maneuver, as well as velocity and trajectory profiles (data from the vehicle of interest, or "ego vehicle").

2. Methods

The present study was designed in order to provide cues about the points raised in the previous section. Participants were required to drive in a monotonous environment. Then, rapidly they had to activate the autonomous mode. Following this, they were informed they could quietly watch a movie on the entertainment screen. They were also informed that at a given moment (without previously knowing when) they would be requested to takeover, and they had to continue manually driving, whatever happens. Though participants were previously trained to the takeover procedure (a single action on a pedal or on the wheel will deactivate the automatic drive and give back control) they were never exposed before the experimental session to the type of event used to test their performance. Thus, we expected them to be surprised, and even more in case of inattention and eventually out-of-the loop state. For this reason, we duplicated the experimental session (same duration, same event). Comparison between the two sessions is expected to give cue about a sort of adaptation to the situation. In order to limit this potential learning which would make comparison between durations difficult, we used independent groups: each participant was exposed to only one duration, twice. To our knowledge this is the first study testing more than two durations (actually four) including both short and long ones. The protocol is detailed in the following sections.

This study was performed in accordance with the principles of the Declaration of Helsinki [43]; all participants received detailed information about the study and gave informed consent. The protocol was approved by a university Bio-ethical committee (CERSTAPS, IRB00012476-2020-15-07-63).

In a static driving simulator located at the Virtual Reality Center facility of Aix-Marseille University, 51 participants were distributed in 4 independent groups (namely G05, G15, G45 and G60) in order to reduce learning effect within the participants, each associated to different automated driving duration: 5 min, 15 min, 45 min or 60 min. They took part in conditional automated driving scenarios divided into two recorded sequences (A and B) with a 10-min break between the two. In each sequence, participants had to react to a takeover request to avoid a simulated road event in front of the driver. During the autonomous driving period, participants were instructed to watch a movie and were informed that the TOR, and therefore the road event, could arise at any time.

2.1. Participants

The participants were students at Aix-Marseille University, aged between 18 and 35 (Mean = 22.1; SD = 3.66). Fifteen (33 %) were females and 36 were males (66 %). All had normal or corrected (by corneal lenses) vision. Recruitment criteria were holding a valid driver's license for at least 6 months or driving accompanied by a qualified driver for at least 2 years (only one participant), not wearing glasses, not being susceptible to simulator sickness (assessed by the Motion Sickness Susceptibility Questionnaire [44]), and having an Epworth scale score of 14 or less (assessing susceptibility to drowsiness [45]).

2.2. Group homogenization

In order to reduce differences between groups, we did not randomly assign participants to a group but grouped them in such a way as to homogenize means and standard deviations across groups regarding age, gender, Epworth Score, and number of kilometers driven per year. Table 1 summarizes mean and standard deviations for each group. From our experience in the field, small differences in age and Epworth Score should not lead to real inter-group heterogeneity. Nevertheless, their potential influence will be discussed later in the article.

Table 1
Summary of participants.

Groups	N	Age	Gender	Kms/year	Epworth
G05	13	22.4 (3.57)	F:4; M: 9	12,800 (7777)	6.5 (3.26)
G15	13	22.7 (4.1)	F:3; M: 10	15,000 (10,100)	9.1 (2.50)
G45	13	21 (2.84)	F:4; M: 9	10,962 (7885.8)	9.9 (3.84)
G60	12	22.6 (4.1)	F:4; M: 8	14.625 (7535.1)	8.75 (3.1)

2.3. Apparatus

2.3.1. Simulator

The static simulator consisted of a triple-screen visual setup (3 video screens, each 24″ landscape format 16/9), a steering wheel from Stellantis Group©, brake and accelerator pedals, and a small screen (10″) used to display the dashboard, located just behind the wheel (see Fig. 1A). The driving environment was generated with SCANeR Studio Software© by AVSimulation©, at a resolution of 1920 × 1080 pixels onto each screen of the triple-screen, providing a 136° horizontal forward field of view. A rearview mirror was simulated in the central front screen, providing a 60° rear field of view, and two side mirrors were simulated respectively on the left and right screens. Two speakers were placed under the left and right screens, while a subwoofer behind the simulator provided simulated engine, road, and traffic sounds. A GoPro Hero7 camera located above the central screen (simulating the driver's front view) recorded driver behavior in order to rate the driver drowsiness level in post hoc analysis.

2.3.2. Driving environment parameters

For both recordings (sequences A and B), the road type was a 2 × 2 highway (See Fig. 1B). The simulated landscape around the highway was relatively featureless, exposing participants to monotonous driving conditions.

Traffic conditions were generated using the swarm generation function of SCANeR Studio Software© to create streaming traffic with the cars in randomly generated positions (range defined in parameters) around the ego vehicle - defined as the queen of the swarm. Simulated weather conditions were a sunny sky, and the experimental room had softened lights and a constant temperature (20° ± 2 Celsius).

Automation was provided by SCANeR Studio Software© according to the current definition of a level 3 highly automated driving system [1]. When the autonomous function was activated, the system was able to safely manage lateral trajectories and longitudinal and lateral speeds, to follow the course of the road, and to keep a safe distance from vehicles ahead (time to collision >2.50 s). The speed during the level 3 automated period was set to 113 km/h, while in manual driving a warning message appeared when the speed reached 120 km/h. The automated driving system was intentionally not programmed to perform any overtaking maneuver. We want to make clear that we choose to conduct the experiment without activating any ADAS brake system at the time of takeover, for two main reasons: firstly, since we needed to identify appropriate/inappropriate manual maneuver after TOR and take-over failure, and manual control trajectory were used as performance indicators, using ADAS (e.g. automatic braking/collision avoidance) would have altered our capability to distinguish failure from success. Secondly ADAS for automatic collision avoidance could induce a bias in the driver's behavioral response.

2.3.3. Non-driving-related task

To ensure that participants did not anticipate the event and were involved in fairly uniform behavior at the time of the TOR, and to heighten their OOTL state, all participants were involved in a non-driving secondary task: watching a movie. This was intended to keep them busy during the session, since doing nothing for up to 1 h is unrealistic. They had to watch a movie (a blockbuster called "Aquaman"; 2018, WarnerBros ©) aimed at attracting and holding viewers' attention [46].

2.3.4. Experimental procedure

Before starting, the participants were allowed to adjust their seats to the most comfortable position. Once in position, the participants began the experimental procedure, consisting of three distinct phases. The first phase aimed to familiarize the participant with the simulator's controls, the simulated environment, the driving task and the TOR procedure. The two other phases were experimental sequences A and B, performed under similar conditions (see below).

Before this first phase (familiarization), participants were instructed how to activate and deactivate the autonomous function. To activate the autonomous function, the driver had to apply the command generally used to activate the headlights. To deactivate the function, the driver had to use either the brake or the accelerator pedal. Furthermore, they were informed that in any situation, the goal was to avoid colliding with other vehicles or the obstacle that might appear at the time of the TOR, and to pursue the course of the

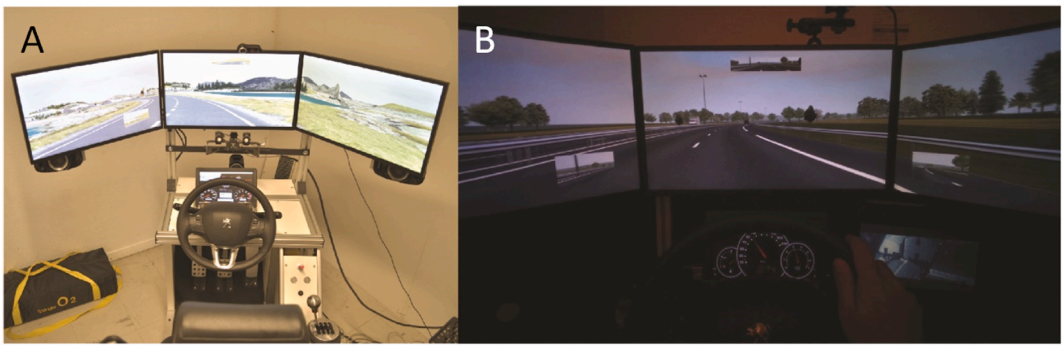

Fig. 1. A. Top view of the static simulator. B. Perspective of a participant in the static simulator.

drive.

2.3.4.1. Familiarization phase. The familiarization phase was designed to ensure a proper understanding of the design of the autonomous function, the way to activate and deactivate it, and the takeover procedure, as well as to accustom the participant to the driving simulation provided by the setup.

In a semi-urban environment, the participants had to follow a lead car during a manual driving period of approximately 6 min. Then the system sent a Manual-to-Autonomous (MtA) notification requiring the participants to activate the autonomous function. Thirty seconds after this MtA driving transition, the system displayed the first TOR involving a critical event with a takeover time (TOT) of 4.95 s. This was the sudden appearance of a police car stopped in the right lane and a car in the security lane, with two surrounding traffic cones. After avoiding this first event (no crashes recorded among the participants), participants had to drive manually for 90 s. After a second autonomous period of about 3 min and 30 s, there was another TOR involving a critical event (TOT = 3 s): a pick-up car located in the middle of a crossroads (no crashes recorded for any participants). Then participants had again to drive manually for one more minute, thus reaching the end of the familiarization phase. Both time budgets were critical, aimed at inducing a high level of expectation about the criticality and uncertainty of TOR events in the subsequent sequences.

2.3.4.2. Main recording sequences. Since the two experimental sequences (A and B) were performed under the same conditions, we only describe here one of them. The simulated environment was a highway ring. The drivers were in a stationary position in the middle of the traffic and they had to reach the speed limit while keeping a safe distance from the lead car (2 security-lanes), overtaking being forbidden. After a period of manual driving (2 min), the system sent a MtA request. After turning on the autonomous driving function, the participants were instructed to watch a movie on the screen located to the right of the steering wheel. They were allowed only to manage the sound level; no timestamp was displayed, to avoid participants counting the time and to prevent any anticipation. At the end of the autonomous period (lasting 5, 15, 45 or 60 min depending on the group), the system sent a TOR notification. The critical event involved here differed somewhat from those used in the familiarization phase. For the experimental sequences, the front car suddenly started braking from 113 km/h to 33 km/h in 5.5 s, providing a time budget of 8.3 s before collision at constant speed (see Fig. 2A and B; all traffic other than the event vehicle was removed). This time budget is recommended by several authors [38,39] as sufficient to ensure safe return to manual control of the car. A participant who did not detect the TOR notification would collide with the event vehicle after the 8.3 s, ending the session. Thus we consider that a successful trial means the driver was able to avoid collision while staying on the road, by either braking and changing lane, or just avoiding the obstacle by changing lane (see Fig. 2C), no specific instruction about what to do was given. A participant successfully deploying an avoidance maneuver would drive manually for 1 more minute.

After the first sequence (A), participants had a 10-min break, left the simulator to stretch their legs, then went back for the second sequence (B). The only parameter differing from sequence A was the starting point of the movie (participants watched the movie from the point where they had stopped it in sequence A). The only thing that the participant knew about the task in both experimental sequences was that at some point there would be a TOR notification involving a driving event that might be avoided, before driving on manually. These instructions purposely created ambiguity and uncertainty for the participant, aiming to simulate an “ecological" use of

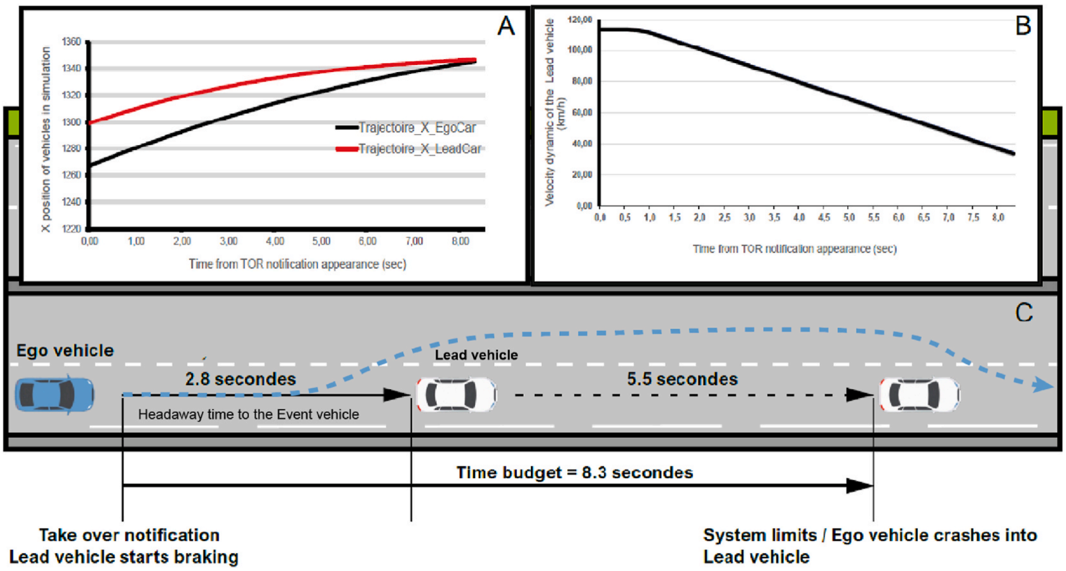

Fig. 2. Schematic representation of takeover event. A illustrates the evolving positions of the lead vehicle (red line) and the ego vehicle (black line) from TOR until crash. B illustrates the deceleration dynamics of the lead car from TOR on. C represents the dynamics of the event at the time of TOR. The blue arrow represents the expected trajectory after take-over and under manual control.

the level 3 automated driving function.

2.3.5. Takeover performance assessment

In order to assess driver behavior at the time of TOR, we measured several dependent variables described in Table 2. The participants were equipped with Pro 2© Tobii Glasses to record the oculomotor response time as the delay between TOR notification and the first eye on the road.

Pre-processing of the oculomotor data (sequencing of oculomotor recording, gaze position mapping on AOIs) was performed with the dedicated software Tobii Pro Lab; manual mapping and detection checks were conducted to minimize software errors. Data analyses were conducted with custom-made Python 3.6, R Studio scripts, and Tobii Pro Lab. For every trial where the takeover maneuver was successful, six variables were extracted from kinematic signals (trajectory and speed) to evaluate the quality of the avoidance maneuver and resumption of control, for each duration group [47].

2.3.6. Assessment of drowsiness over time

The level of drowsiness was determined according to a method proposed by Wierville and Ellsworth [48] and based on subjective assessment by video analysis. We used the Observer Rating of Drowsiness (ORD) scale [49], with levels of drowsiness numbered (level 0: alert state, to level 4: extremely drowsy) for each minute of the driving sequences and for each subject recorded. Due to the general reliability of paired ratings by the same rater, there was only one rating per recording.

2.3.7. Statistical analysis

Since prior tests did not satisfy normality and/or variance homogeneity, a Kruskal-Wallis test was used to compare groups within a sequence, both on times and manual-driving-related measures. With grouped data, a Mann-Whitney test was applied when necessary, followed by a post-hoc Kruskal-Nemenyi test to identify the difference. Finally, comparisons between sequences A and B were performed using a Wilcoxon test. All results are expressed and presented in figures as mean ± standard deviation.

3. Results

Several indicators were measured to assess driver performance: first, crash rates and drowsiness levels, followed by takeover times between TOR and the resumption of manual driving, and finally, manual driving performance in managing the event.

3.1. Crash rates by autonomous driving duration

Of the recorded driving sequences, 14 % ended in a crash. As shown by group in Fig. 3A, this corresponds to a crash rate of around 15 % for G05 in both sequences A and B, 23 % for G45 in sequence A and 31 % in sequence B, 8 % for G60 in sequence A and 15 % in sequence B. No crash was observed for G15 in either sequence. Trials ending in a crash were excluded from subsequent analyses but will be discussed at the end of the article. None of the participants crashed twice. The crash rate (Fig. 3A) was not linearly correlated to the duration of autonomous driving, around 15 % of crashes being observed for the 5-min conditions and none for the 15-min conditions. Moreover, crash rates for the 45-min conditions were higher on average than for both the longer and shorter conditions.

3.2. Drowsiness-level ratings over time

The mean maximum level of ORD increased with duration, as shown in Fig. 3B, and the mean dynamic of ORD shown in Fig. 3C and D, respectively for group G45 and group G60, supporting a linear dynamic for the emergence of drowsiness with prolonged autonomous driving. This might suggest that the effect of duration on takeover performance, if primarily due to drowsiness, should also follow a linear dynamic. The following results will show it is not the case. Below, the performance of participants who succeeded in avoiding the crash is analyzed, with takeover performance measured first in terms of time from TOR to takeover, then in terms of

Table 2
Summary of dependent variables.

Dependent Variable	Explanation
Takeover time (s) (TOT)	Time between TOR and start of manual driving
First eye on road (s) (FER)	Time between TOR and the first gaze on driving area
Decision Time (s) (DT)	Time between the first gaze on driving area and start of manual driving
Minimum Time-to-collision (s) (Min TTC)	Shortest time before crashing into lead vehicle, measured just before the lane change.
Max Deceleration ($m.s^{-2}$) (MD)	Maximum value of deceleration profile measured on the period starting when level 3 function is turned off until min velocity measured.
Time-to-reach maximum deviation (s) (TMD)	Time needed to reach max lateral deviation during maneuver starting at 0.75 m from its initial position.
Time from maximum deviation to right lane position (s) (TMR)	Time elapsed since the moment of maximum deviation to return at least 0.5 m from the initial position in the right lane.
Distance from maximum deviation to right lane position (m) (DMR)	Distance covered from the moment of maximum deviation to return at least 0.5 m from the initial position in the right lane.
Speed variation ($km.h^{-1}$)	Profile of speed (difference between velocity and initial velocity at TOR)
Lateral deviation (m)	Lateral distance from initial position at TOR.

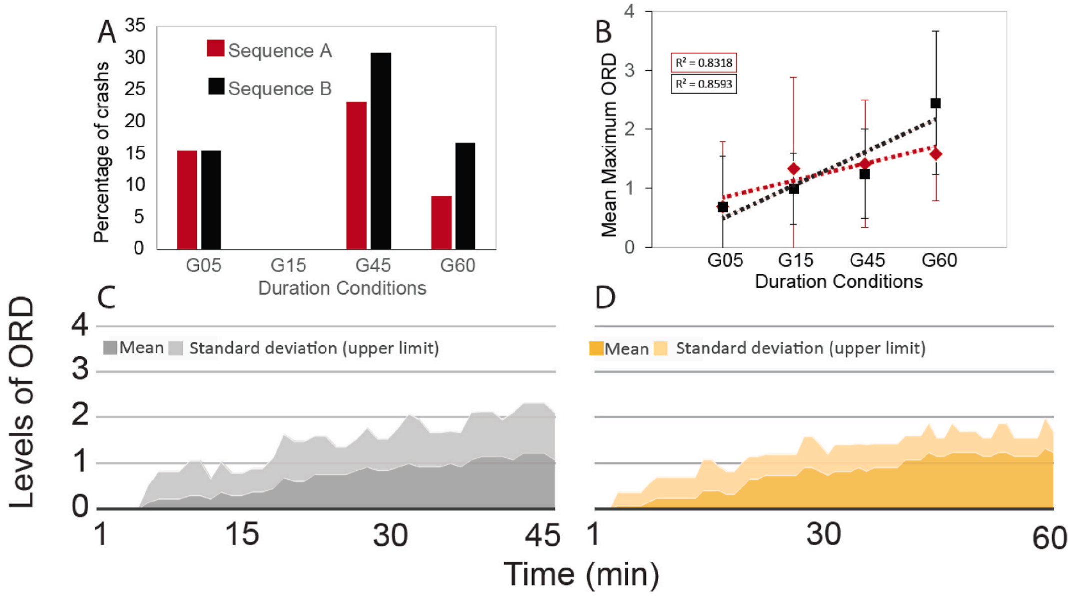

Fig. 3. Percentage of crashes observed for each group and each sequence (Panel A). Mean maximum ORD rated for each group in sequence A (red) and sequence B (black, Panel B). Bottom panels illustrate mean assessment of ORD, for group G45 (Panel C) and G60 (Panel D). Dark areas in both panels illustrate average values while light areas illustrate maximum standard deviation.

resumption of manual control of the car.

3.3. *Time to takeover*

After the four groups' sequences were compared, the data was pooled to repeat the intra-sequence and inter-sequence comparisons.

3.3.1. *Impact of duration of autonomous driving*

3.3.1.1. First drive (sequence A). The left panel of Fig. 4 shows the time taken to return to the loop by participants in sequence A. The sum of the time to FER (Fig. 4A) plus the time between this first saccade and the decision to turn off the autonomous function (Fig. 4B) is considered as the takeover time.

The three Kruskal-Wallis tests conducted for each parameter did not reveal any statistical difference between groups (see Table 3, all p-values >0.05). Time between the TOR and the first saccade toward the road scene did not significantly change with autonomous driving duration (Fig. 4A). However, despite the absence of significant difference, a positive linear trend was observed for decision time: the time taken to decide to deactivate the autonomous function increased with autonomous driving duration (from 3.21 ± 1.40 s for G05 to 4.51 ± 1.34 s for G60, see Table 3). Consequently, mean takeover time (Fig. 4C) followed a similar trend (see Table 3).

3.3.1.2. Second drive (sequence B). Like sequence A, the three Kruskal-Wallis tests conducted did not reveal any statistical difference between groups (see Table 3, all p-values >0.05) and visual inspection of distribution of FER times did not reveal any effect of duration (Fig. 4D). Unlike sequence A, however, decision time (Fig. 4E) did not follow a linear trend. Decision times were shorter for G05, but similar for G15 and G45, and longer for G60. A positive linear trend was observed (as with sequence A) for mean takeover times (Fig. 4F; from 3.22 ± 1.54 s for G05 to 4.18 ± 2.15 s for G60, see Table 3).

3.3.1.3. Comparison between sequence A and sequence B. Comparisons between sequences A and B for each group showed no difference (Wilcoxon test: $p > 0.05$): the groups' times remained fairly stable from sequence A to sequence B (see Fig. 4C and F).

3.3.2. *Pooled data*

This lack of significance, despite the positive linear trends in takeover times systematically observed, may be explained by the relatively small group size. However, given the trends, we decided to pool the data into two groups of similar autonomous driving duration: one under 30 min (pooling G05 and G15, <30), and one over 30 min (pooling G45 and G60, >30).

3.3.2.1. Sequence A. Comparison between data from Under 30 and Over 30 (see Fig. 5A) showed a significant difference in takeover time between the groups (Fig. 5, left panel). Analysis revealed significantly longer takeover times for >30 than for <30 ($TOT_{<30} = 3.35 \pm 1.33$ s; $TOT_{>30} = 4.33 \pm 1.24$ s; Mann-Whitney test, $p = 0.019$), with a moderate effect size ($r = 0.351$).

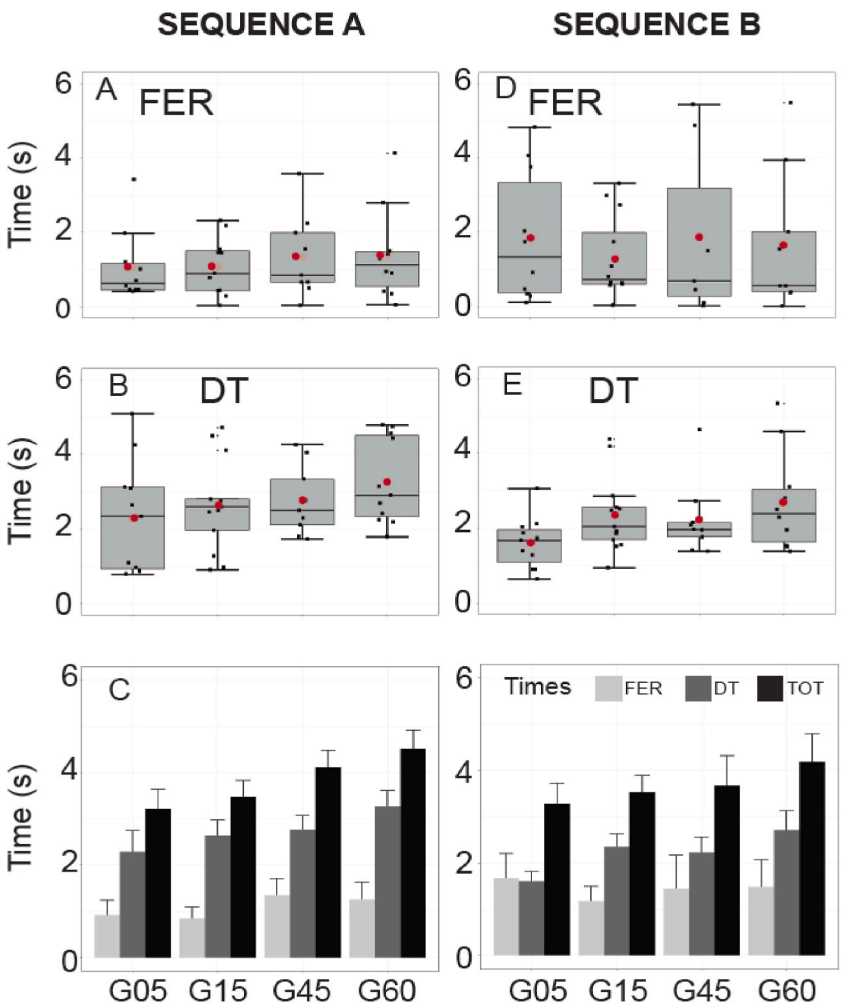

Fig. 4. Reaction times measured for sequence A (left panels, A, B, C) and sequence B (right panels, D, E, F). Panels A, D: boxplot illustrating the distribution of first eye on road time (FER). Panels B, E: boxplot illustrating decision time (DT) showing distribution of measures across groups. Colored dots show the mean for each group in each boxplot, black dots show individual measurements. Horizontal bars at the ends represent standard deviation, central black horizontal lines represent median for each group. Bottom panel (C, F): bar plots showing mean values for each reaction time and each group. Black bar plots illustrate takeover time (TOT) for each group. Horizontal black lines illustrate standard error for each mean value.

Table 3
Time values. Table shows mean time values (standard deviation) for all groups measures in both sequences.

Time	G05	G15	G45	G60	χ^2	p
FER (s) Sequence A	0.92 (1.02)	0.83 (0.90)	1.34 (1.11)	1.25 (1.25)	1.55	0.671
Sequence B	1.67 (1.77)	1.17 (1.15)	1.44 (2.16)	1.48 (1.86)	0.60	0.895
DT (s) Sequence A	2.24 (1.48)	2.56 (1.23)	2.76 (0.93)	3.27 (1.14)	3.17	0.366
Sequence B	1.60 (0.69)	2.35 (1.00)	2.22 (0.99)	2.70 (1.33)	6.16	0.104
TOT (s) Sequence A	3.21 (1.40)	3.47 (1.31)	4.07 (1.10)	4.51 (1.34)	6.31	0.097
Sequence B	3.27 (1.51)	3.52 (1.34)	3.67 (1.94)	4.18 (1.92)	1.40	0.703

3.3.2.2. Sequence B. In sequence B, visual inspection (Fig. 5B) showed a similar pattern, with quite similar FER and slightly greater DT and TOT for the Over 30 min group. Statistical analysis did not reveal any significant difference in the three reaction times between Under 30 min and Over 30 min. Subsequent analysis revealed that the absence of statistical difference (compared to sequence A) for DT is due to the high standard deviation measured for the >30 and the more similar means of each group ($TOT_{<30} = 3.41 \pm 1.40$s; $TOT_{>30} = 3.94 \pm 1.90$s).

3.3.2.3. Comparison between sequence A and sequence B. An inter-sequence pairwise comparison on times for each group revealed that

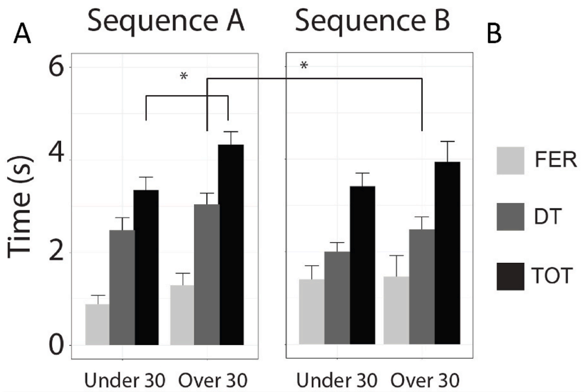

Fig. 5. All times for groups Under 30 min and Over 30 min in both sequences, Panel A for sequence A, Panel B for sequence B. Bar plots show mean values for each time and each group. Significant differences are indicated by a star ($p < 0.05$). Abbreviations are defined on Table 2.

DT was significantly shorter in sequence B than in sequence A for the Over 30 min group ($Ov30_{dtA} = 3.04 \pm 1.05s$); $Ov30_{dtB} = 2.48 \pm 1.17s$); Wilcoxon test: $p < 0.05$).

3.4. *Avoidance manoeuvre and quality of control*

Analyses are focused here on manual driving performance after takeover, based on the driving indicators described in 2.3.5.

3.4.1. *Influence of autonomous driving duration*

As shown in the panels A,C,E and G of Fig. 6, lateral deviations to manage the situation varied widely in amplitude in all groups for sequence A. Inter-individual variability was mostly due to variable RTs, also some drivers braked before changing lane, others did not (see speed variability on Fig. 6B, D, 6F and 6H respectively for group G05, G15, G45 and G60). Consequently, deviations occurred at different times and some participants produced more than one deviation. Thus there are several peaks in the average curves, the highest at 10 s means this is the most probable time when an initial deviation occurred. A successful reaction (no crash) means necessarily the driver changes lane. Indeed, as stated in the methods section, each lane width is 3.5 m. Thus a lateral deviation of at least 3 m means a lane change (occurring in average 10 s after takeover).

Interestingly, according to the secondary hypothesis (see section 1.3), this variability was reduced in sequence B, with smaller standard deviations reflecting a more homogeneous control of the car during the maneuver in all groups. The velocity profile, illustrated in the right panels of Fig. 6 as a speed variation profile, showed homogeneous average profiles but large standard deviations in sequence A. Standard deviations and mean profiles in sequence B varied less between groups, reflecting a more homogeneous participant response.

On no indicator (see 2.3.5) did distributions respect the normality assumption (Shapiro-Wilk test, $p < 0.05$), and therefore Kruskal-Wallis tests were used for comparisons between groups in sequence A and sequence B. Of the five indicators (see Table 4), only two (TMR and DMR) showed a statistically significant difference relative to duration (see Fig. 7), in sequence A only (none in sequence B). The time taken to return to the initial position in the right lane after reaching the maximum deviation (TMR) was influenced by the duration of the autonomous drive (Kruskal-Wallis test; $p = 0.02$; $\eta^2 = 0.159$ (large effect size)). Post-hoc Kruskal-Nemenyi tests indicated a significantly longer time for group G05 ($M_{TMR} = 8.07 \pm 2.93$ s) than for group G15 ($M_{TMR} = 4.40 \pm 2.26$ s) ($p = 0.017$) (Fig. 7A).

In line with the previous results, a significant difference between groups was observed in sequence A (Kruskal-Wallis test; $p = 0.01$; $\eta^2 = 0.184$ (large effect size)) on the distance covered by drivers from the maximum deviation position until they returned to the initial right lane position (DMR). Post-hoc Kruskal-Nemenyi tests revealed ($p = 0.022$) a significantly longer distance for group G05 ($M_{DMR} = 214.35 \pm 100.55$ m) than for group G15 ($M_{DMR} = 100.53 \pm 63.40$ m) (Fig. 7B). The last three indicators, which did not show any difference between the initial four autonomous driving duration groups, were merged, as previously for times, into two groups: Under 30 min and Over 30 min. When data were compared between these two groups, no statistical difference was observed (Wilcoxon test; $p > 0.05$). The following section thus presents inter-sequence comparisons with all data pooled.

3.4.2. *Comparison between sequences A and B*

Analyses of lateral deviation and speed differential profiles (see Fig. 6) showed that, in all groups, behaviors at takeover differed between sequences A and B. In sequence A, some participants perceived the TOR earlier, quickly braked to deactivate the function, but took some time before starting the avoidance maneuver and changing lanes. Others braked sharply on detecting the TOR, then waited briefly before overtaking the lead car. In other words, there was a wide range of takeover behaviors in all our groups in sequence A. However, behaviors differed less in sequence B, with a smaller standard deviation in all groups. In the light of the major behavioral discrepancies between sequence A and sequence B, we conducted an inter-sequence comparison (Fig. 8).

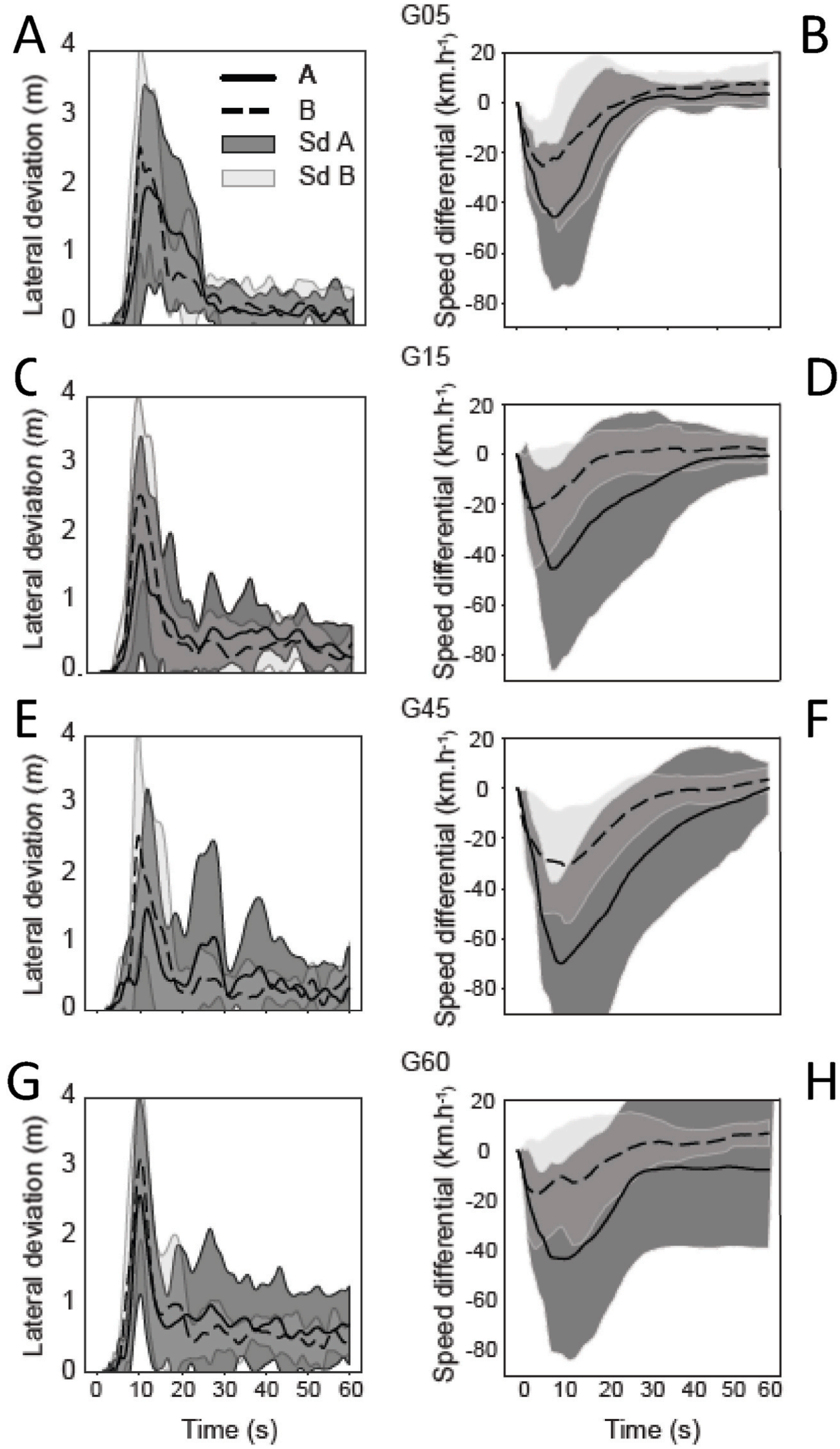

Fig. 6. Average lateral deviation from initial lateral position of the car at the start of manual driving (left panels: A, C, E, and G) and average speed differential from start of manual driving (right panels: B, D, F and H) for each group (from top to bottom: G05, G15; G45 and G60). Deactivation of automatic driving occurs at time = 0 s. Black lines illustrate mean profiles for sequence A (grey areas = sd) while dashed black lines illustrate mean profiles for sequence B (shaded grey areas = sd).

3.4.2.1. Maximum deceleration (MD). The distribution of the values (Fig. 8A) revealed two different performance profiles in both sequences, with one set of participants presenting deceleration values above -5 m s^{-2} and another presenting values below -5 m s^{-2}. A Mann-Whitney test conducted on maximum deceleration revealed a significant difference between sequences A and B (MD_{SA} =

Table 4
Ego car measurements taken into account to assess takeover performance during maneuver. Each cell pools mean results (standard deviation in parenthesis) for sequence A (top of the cell) and sequence B (bottom of the cell).

Time	G05	G15	G45	G60	χ^2	*p*
Min TTC (s) SEQUENCE A SEQUENCE B	0.290 (3.88)	−0.258 (7.77)	−3.389 (8.46)	1.09 (5.94)	3.01	0.39
	2.54 (4.08)	3.62 (1.86)	2.55 (2.50)	4.09 (3.71)	4.04	0.25
MD ($m.s^{-2}$) SEQUENCE A SEQUENCE B	−9.36 (3.51)	−6.26 (4.64)	−9.13 (4.10)	−6.08 (4.46)	6.12	0.1
	−7.01 (4.73)	−5.23 (4.69)	−5.67 (4.73)	−4.02 (3.67)	3.38	0.33
TMD (s) SEQUENCE A SEQUENCE B	9.92 (3.94)	10.26 (7.91)	13.88 (9.18)	9.50 (7.19)	3.14	0.36
	7.69 (4.27)	6.68 (2.07)	7.32 (2.88)	6.53 (3.97)	1.71	0.63
TMR (s) SEQUENCE A SEQUENCE B	8.07 (2.93)	4.40 (2.26)	6.12 (2.85)	8.93 (6.77)	9.37	**0.02**
	6.23 (3.44)	6.89 (7.77)	5.61 (3.33)	7.25 (5.02)	1.90	0.59
DMR (m) SEQUENCE A SEQUENCE B	214.35 (100.55)	100.53 (63.40)	108.37 (57.30)	223.92 (178.76)	10.35	**0.01**
	176.77 (107.17)	202.81 (253.24	199.62 (154.78)	122.54 (65.95)	2.052	0.56

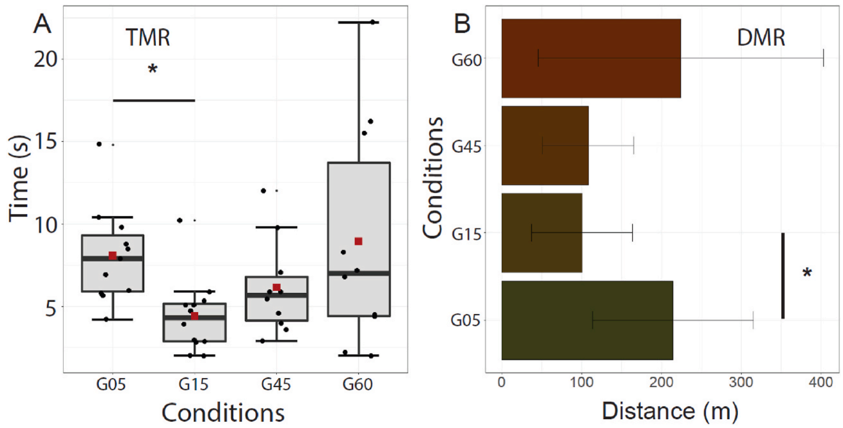

Fig. 7. Illustration of time (Panel A, TMR = Time from Maximum deviation to Right lane position) and distance (Panel B, DMR = Distance from Maximum deviation to Right lane position) to return to initial right lane position after reaching maximum lateral deviation during sequence A. Panel A: colored squares show the mean for each group in each boxplot, black dots show individual measurements. Horizontal bars at the ends represent standard deviation, central black horizontal lines represent median for each group. Panel B: Horizontal black lines for each bar plot illustrate standard deviation. Significant differences are indicated by a star (p < 0.05).

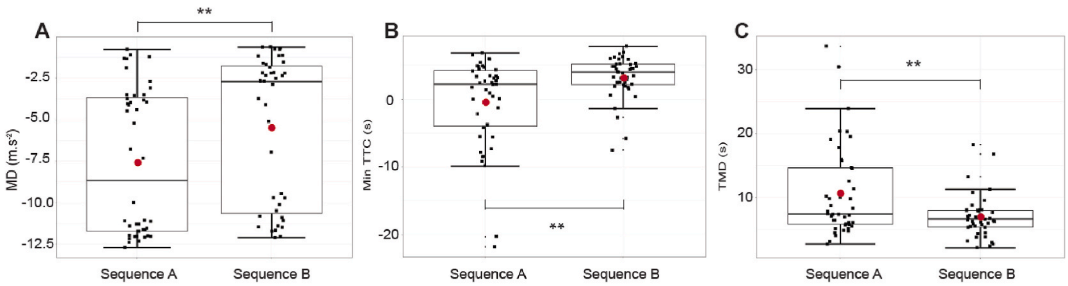

Fig. 8. Comparison of Maximum deceleration (Panel A), Minimum TTC (Panel B), and Time to reach maximum deviation (Panel C) between sequence A and sequence B. Red dots show mean value for each measure. Significant differences are indicated by double stars (p < 0.01). Abbreviations are defined on Table 2.

-7.58 ± 4.36 m/s^2); $MD_{SB} = -5.40 \pm 4.46$ m/s^2); $p < 0.01$), with a moderate effect size ($r = 0.320$). This difference indicates that, in addition to showing two different speed profiles during the maneuver, drivers tended to slow down less during the maneuver in sequence B than in sequence A.

3.4.2.2. Minimum time to collision (min TTC). The Min TTC (Fig. 8B) largely differed between sequences A and B. The Mann-Whitney test revealed a significant inter-sequence difference ($p < 0.01$; with a moderate effect size $r = 0.320$). In sequence A, an average negative value was observed, illustrating a late maneuver after resuming control of the vehicle. In sequence B, Min TTC revealed that on average drivers managed the event at higher speeds than in sequence A (Min $TTC_{SA} = -0.47 \pm 6.63$ s; Min $TTC_{SB} = 3.32 \pm 3.11$ s).

3.4.2.3. Time to reach maximum deviation (TMD). As with the previous variables, the time taken to reach maximum deviation differed significantly between sequences A and B (Fig. 8C, $TMD_{SA} = 11.24 \pm 7.04$ s; $TMD_{SB} = 6.99 \pm 3.36$ s); $p < 0.01$, with a large effect size ($r = 0.763$). Thus drivers did their avoidance maneuver faster in sequence B than in sequence A.

4. Discussion

Previous studies have shown that exposure to prolonged autonomous driving weakens vehicle control during avoidance manoeuvres [42], reduces the ability to react to a sudden and unexpected event [20], and increases the likelihood of drowsiness, attentional shifting, or mind wandering that could induce a loss of situation awareness and an out-the-loop state [50]. However, to our knowledge, none of these previous studies applied a larger range of durations of autonomous driving to determine their influence on takeover performance. We therefore investigated a gradient of four intervals of autonomous driving, assessing subsequent takeover performance through crash rate, time taken to regain manual control, and driving behavior following the TOR.

4.1. Impact of autonomous driving duration on takeover performance

Drowsiness (which impairs driving ability [19]) is known to increase linearly with increasing driving durations, as confirmed here by ORD ratings that increased progressively with increasing duration of autonomous driving. Few participants reached the maximum ORD level, and consequently, few of them were extremely drowsy at the time of TOR. Beyond the obvious safety issue of drowsiness, these results reveal heterogeneity in the dynamics governing participant drowsiness. Video examination of the state, during the last minute before TOR, of participants who crashed shows that only two crashed because of extreme drowsiness (level 4 on the Wierville and Ellsworth scale [48]), one in G05, one in G45 and interestingly, only in sequence A. Other participants were completely absorbed in the movie and did not detect and react to the TOR.

Strictly speaking, the assumption of linearity should have led to the observation of a greater number of collisions with longer autonomous phases, with a maximum ORD at the time of TOR for the longer duration groups. However, the crash rate was not correlated with the mean maximum ORD. Even though the maximum ORD was often reported in the longer duration groups, it did not appear systematically at the critical moment of TOR. The crash rate was highest in G45, followed by G60, but there were also some crashes in the shorter duration groups in both sequences (see Fig. 4A). All these observations are in contradiction with the supposed linear influence of duration of autonomous driving on takeover performance and safety. This highlights the weight and the role of the NDRT and individual factors, like fatigue due to lack of sleep, which will be discussed in the following sections.

4.1.1. Are reaction times an indicator of performance?

The experimental design involving four different durations of autonomous driving did not enable us to establish a clear relationship between duration and reaction times. The positive linear trend in takeover times needs to be considered in association with the crash rate, which was not correlated to the initial duration gradient: crashes were observed even in the shortest duration group. This suggests that poor performance is not necessarily due to the duration of autonomous driving, but rather to other factors.

Results from pooled data show that the mean takeover time (from TOR to action) was impacted by the duration of the autonomous drive. Exposure to more than 30 min in level 3 autonomous driving increased the time taken to deactivate the function and regain in-the-loop control (manual driving). This confirms and reinforces previous findings by Bourelly et al. [42], on the effect of prolonged autonomous driving on time taken to react to a TOR. They reported an increase of 0.5 s in takeover time for a 60-min duration relative to a 10-min period (paired groups), while we measured a mean difference greater than 1 s between takeover times for durations of under 30 min and those for durations of over 30 min. According to the literature on factors influencing times of action in driving tasks, an increase of 1 s is of critical concern from an ecological perspective. The absence of effect of autonomous driving duration on times in sequence B using the pooled data may be explained by the change in decision time between the two sequences for the over 30 min group. Decision time became quicker in sequence B and closer to that measured for the under 30 min group in sequence B, likely due to better decision-making when faced with the event. Thus, while the times measured here suggest some influence from the duration of autonomous driving, they are not necessarily informative about performance. These times should be considered in the light of the time budget of the takeover situation. Autonomous driving duration likely interacts with other factors inherent to the driver or the environment to determine performance.

4.1.2. Indicators of manual driving performance

In all groups manual driving performance varied more in sequence A than in sequence B, with great heterogeneity in the management of the event after takeover. These observations can be compared with those reported by Volgelpohl et al. [20], who found a strong difference between fatigued and in-control (alert) drivers in terms of avoidance performance. Accordingly, the fact that some drivers may have been more fatigued than others may explain the wide spectrum of profiles in our population. Unexpectedly, the only difference measured in manual driving performance was between the two shorter duration groups, which could also indicate the influence of factors unrelated to duration (i.e., degree of complacency with the system, effect of the NDRT, or even expectations regarding the scenario).

The repetition of the takeover event impacted driving behaviors for all groups in sequence B. This interesting result can be compared to Volgelpohl et al. who found reduced variability in control after takeover (Volgelpohl et al. [20]). The time taken to start the avoidance maneuver reduced, and velocity increased. Although from an ecological perspective, high velocity could be considered dangerous, here the drivers' higher velocity profiles in sequence B reflect better decision-making and confidence in their ability to manage the situation, as shown from the differences in decision time between sequences A and B. We found fairly consistent heterogeneity of performance across groups, with similar behavioral variation among participants in long and short duration groups. This indicates the non-linearity of the effect of autonomous driving duration, the complex interactions in this context, and strong idiosyncrasies. Our main outcomes are similar to observations reported by Kuehn et al. [37]: they attributed the wide range of takeover maneuvers they observed to interaction effects of the NDRT, driver's attributes, and type of takeover situation used.

Taken together, crash rates, reaction times, and manual driving performance indicators point to a likely heterogeneity in the use of these new autonomous functions. The 15-min autonomous driving period appears to induce the most homogeneous performances, with no crashes and good driver performance. On this basis, limiting autonomous driving to short periods, say between 15 and 30 min, might be sensible. However, we need to consider how far the training, endogenous factors such as driver state, or the driving scenario may have been responsible for the high inter-individual variability reported here, which clearly affected the statistical significance of our results.

4.2. Influence of the non-driving-related task

The NDRT we used may have influenced takeover performance: a superhero action movie, "Aquaman" (2018, WarnerBros ©). It was chosen because it was long enough to be watched over the two driving sequences (especially given groups G45 and G60). This blockbuster follows a specific format alternating action phases and calmer periods. One of the most important aspects of an NDRT is what Jarosch et al. [51] defined as the way it affects the driver's arousal level, and thus how it affects cognitive transition (the return to being "On the loop" of control). This is especially important when the participant's attention is fully devoted to the task, as with the movie in the present study. Du et al. [52] demonstrated that performance at takeover is correlated to the emotional state of the driver, reporting that positive emotions induce a shorter takeover time, whereas level of arousal has no influence. This leads us to suggest that watching this type of movie could, depending on the scene the driver is watching at TOR, trigger either a positive or a negative emotion (as described by Russel [53]) and thereby modulate the individual's takeover performance. Indeed, the scenes our participants were watching just prior to the TOR differed substantially between groups and sequences. The content varied in terms of temporality, dynamics, and subjects, with slow scenes, suspenseful scenes, several scenes consisting of dialog, and fight scenes. In G15 (no crashes), the scenes watched during the last minute before the TOR were fight scenes containing several high intensity events, in both sequences. In G45's sequence A (crash rate of 23 %), the scene watched just before the TOR was a dialog scene with no action but revealing an important plot event, whereas in sequence B (crash rate of 31 %), it was a fight scene. The same pattern was observed for G60: a dialog scene in sequence A and a fight scene in sequence B, with more accidents in B than in A. Moreover, G05, where no fight scene was watched just prior to TOR, but a storytelling scene in sequence A and a suspenseful scene in sequence B, showed a consistent crash rate (15 %) between sequences. Such differences in the type of scenes being watched, with the unknown emotional valence induced by these scenes, may add to the idiosyncrasies in takeover performance. Further, the task "watching a movie" could have been approached differently by our participants, depending on their personal interest in this kind of movie and their perception of movie-watching itself: for some, it may have been a monotonous or boring task, whereas for others it was an exciting task.

Actually, any secondary task with unpredictable consequences in terms of workload and/or emotional effect on the driver could constitute a major safety problem. In other words, the assumption that the same movie will induce exactly the same reactions in all drivers is risky. More studies investigating the emotional impact and dynamics of non-driving-related tasks are needed to better define and control the factors which really determine the return to the cognitive loop of control.

4.3. Effect of repetition

The procedure used here revealed that repetition reduced variability, likely due to a process of adaptation or learning by participants. As reported by previous studies showing an improved return to the manual control loop [47], we found a clear difference in performance in sequence B for participants who had already successfully resumed control of the vehicle, as well as a better understanding of the situation. These observations are supported by the difference in control resumption behaviors between sequence A and sequence B. In sequence A, some participants detected the TOR, resumed control of the vehicle by strongly braking (high maximum deceleration), then waited a little before executing the overtaking maneuver (longer decision time and minimum time-to-collision exceeding the time budget). In sequence B, velocity profiles showed that participants resumed control of the vehicle by lightly braking and immediately changing lane at high speed. This evolution is in line with Gold et al. [38], who reported that the longer drivers had to decide what to do, the less they used their brakes (see Fig. 6D). This pattern of behavior suggests a potential adaptation and transfer of what they learned from prior exposure [54,55] to the same situation, which is also consistent with the difference observed in decision time between sequence A and B.

While for some participants, repetition induced positive effects on performance, for others it was the opposite (crash rates remaining fairly stable between the two sequences), highlighting the irrelevance of non-supervised learning regarding overall safety. We join Sibi et al. [56] in calling for further investigation on how repetitions modulate mental models and behavioral responses.

4.4. Limitations

Certain limitations should be taken into account when interpreting the present results. The first is the level of expertise and the age of our participants with regard to the autonomy level studied here, likely to be implemented in cars purchased by older people. Being less experienced in using high-level driving aids or new assisted driving technologies could increase variability in behavioral responses.

Second, we identified two issues concerning the non-driving-related task. We did not control the emotional valence for individual participants of the movie scenes watched just prior to takeover. Moreover, watching a film during a long trip might not be a familiar task to all subjects, underlining their motivational heterogeneity.

Though it is now highly recommended that all new vehicles must be equipped with specific ADAS to avoid (or limit) the risk of front collision we decided to shut down all ADAS during the experimental sessions. This choice resulted from the compromising conflict between experimental control and ecological validity. In addition, it may be argued that such systems could fail, and even if these situations may rarely occur, it is known that it can result in fatal accidents.

Finally, the use of a simple driving simulator (instead of a more sophisticated, immersive one), or even a real car on a real road, may be criticized. Nonetheless, it is also well-known that simulator studies are particularly relevant to study human behavior. According to the topic of research, even small and static platforms allow to reproduce representative behavior. Similar simulators were used to evaluate human behavior after TOR [57,58]. Furthermore, it has been shown that the impact of driving duration and sleep deprivation can be assessed through real car driving or fixed base low-cost simulator in a comparable way [59]. In addition, it is also well known that simulator studies emphasize driving impairments noticeably regarding attention and drowsiness consequences.

Obviously, our results cannot be directly transferred to more ecological situations without confirming them with road studies. Drivers will probably be more prudent when driving a real car, while exposed to a real danger. Even though some recent examples illustrate that such system failures can occur in real life conditions.

From a more general point of view, the question of transfer between virtual reality (simulators) and real life is a long debated one (and not completely closed). It remains that simulation is a safer and more reproductible way to experimentally study human behavior [60–62].

5. Conclusion

In conclusion, the present study showed that long duration of autonomous driving is not the main factor influencing the take-over performance. Reaction times related to the take-over maneuver do not necessarily reflect the level of TOR management performance. Trajectory analysis is more relevant. Also, repetition of TORs induced a non-systematic improvement of performance, suggesting a potential need for a specific and systematic training to this situation.

Thus, there is a pressing need to investigate the mechanisms behind driver idiosyncrasies. An increased understanding of their determinants could lead to better guidance on how to prepare the driver for good autonomous driving practices. It might also help to determine the type of features desirable in the human-machine interface and to improve driver monitoring.

Overall, the following aspects should be considered regarding future studies concerning human behavior while driving and using the autonomous level 3 function since they are probable sources of variability: (i) takeover performance is likely modulated by individual features weighted by the duration of the autonomous driving phase, and (ii) non-driving-related tasks have a critical and driver-specific influence.

Data availability statement

Data will be made available on request.

CRediT authorship contribution statement

Arthur Portron: Writing - original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **Gaëtan Perrotte:** Writing - review & editing, Formal analysis. **Guillaume Ollier:** Software, Data curation. **Clément Bougard:** Writing - review & editing, Conceptualization. **Christophe Bourdin:** Writing - review & editing, Resources, Conceptualization. **Jean-Louis Vercher:** Writing - review & editing, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Jean Louis Vercher reports financial support was provided by Fondation MAIF pour la Recherche. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This study was funded by Fondation MAIF pour La Recherche, in collaboration with the OpenLab - Automotive Motion Lab (agreement between Stellantis, Aix-Marseille University and CNRS). The authors would like to thank Dr. Fabrice Sarlegna for his supportive feedbacks, Mr. Philippe Vars from Stellantis group for his technical support at some points of the research.

References

[1] Committee, S.O.R., & others, Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, SAE International: Warrendale, PA, USA, 2021.
[2] Mercedes Benz, The Front Runner in Automated Driving and Safety Technologies, 2022. https://group.mercedes-benz.com/innovation/case/autonomous/drive-pilot-2.html.
[3] March 4 Honda, Honda to Begin Sales of Legend with New Honda SENSING Elite, 2021. https://global.honda/newsroom/news/2021/4210304eng-legend.html.
[4] A.G. Audi, Audi Piloted Driving, 2016. https://www.audi-mediacenter.com/en/audi-at-the-ces-2016-5294/piloted-driving-5303.
[5] Volvo Cars, Intellisafe Autopilot, Intellisafe Autopilot, 2016.
[6] R. Bradley, Tesla Autopilot. MIT Technology Review, https://www.technologyreview.com/s/600772/10-breakthrough-technologies-2016-tesla-autopilot/, 2016.
[7] O. Carsten, M.H. Martens, How can humans understand their automated cars? HMI principles, problems and solutions, Cognit. Technol. Work 21 (2019) 3–20.
[8] J.A. Cameron, A.J. Sottile, Automation effects on a driver's vigilance in the automated highway system, in: Tech. Rep., ume 1, Crew System Ergonomics Information Analysis Center Wright-Patterson AFB OH, 1997. Final report.
[9] N. Merat, A. Jamson, How Do Drivers Behave in a Highly Automated Car? 5th Driving Assessment Conference, 2009, pp. 514–521.
[10] N.B. Sarter, D.D. Woods, Situation awareness: a critical but ill-defined phenomenon, Int. J. Aviat. Psychol. 1 (1) (1991) 45–57.
[11] M.R. Endsley, C.A. Bolstad, Individual differences in pilot situation awareness, Int. J. Aviat. Psychol. 4 (3) (1994) 241–264.
[12] E.R. Boer, M. Hoedemaeker, Modeling driver behavior with different degrees of automation: a hierarchical decision framework of interacting mental models. Proceedings of 17^{th} European Annual Conference on Human Decision Making and Manual Control, 1998, pp. 63–72.
[13] N. Merat, B. Seppelt, T. Louw, et al., The "Out-of-the-Loop" concept in automated driving: proposed definition, measures and implications, Cognit. Technol. Work 21 (2019) 87–98.
[14] M.R. Endsley, Situation awareness, in: G. Salvendy (Ed.), Handbook of Human Factors and Ergonomics, John Wiley & Sons, 2006.
[15] J. Nilsson, P. Falcone, P, J. Vinter, Safe transitions from automated to manual driving using driver controllability estimation, IEEE Trans. Intell. Transport. Syst. 16 (4) (2015) 1806–1816.
[16] D.B. Kaber, M.R. Endsley, Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety, Process Saf. Prog. 16 (1997) 126–131.
[17] B. Berberian, B. Somon, A. Sahai, J. Gouraud, The out-of-the-loop bBrain: a neuroergonomic approach of the human automation interaction, Annu. Rev. Control 44 (2017) 303–315.
[18] J. Horne, L. Reyner, Vehicle accidents related to sleep : a review, Occup. Environ. Med. 56 (5) (1999) 289.
[19] N. Schömig, V. Hargutt, A. Neukum, I. Petermann-Stock, I. Othersen, The interaction between highly automated driving and the development of drowsiness, Procedia Manuf. 3 (2015) 6652–6659.
[20] T. Vogelpohl, M. Kühn, T. Hummel, M. Vollrath, Asleep at the automated wheel. Sleepiness and fatigue during highly automated driving, Accid. Anal. Prev. 126 (2019) 70–84.
[21] N. Merat, D. de Waard, Human factors implications of vehicle automation: current understanding and future directions, Transport. Res. F Traffic Psychol. Behav. (2014) 193–195, 27 Part B.
[22] M. Körber, C. Gold, D. Lechner, K. Bengler, The influence of age on the take-over of vehicle control in highly automated driving, Transport. Res. F Traffic Psychol. Behav. 39 (2016) 19–32.
[23] J. Gong, X. Guo, L. Pan, C. Qi, Y. Wang, Impact of age on takeover behavior in automated driving in complex traffic situations: a case study of Beijing, China, Sustainability 14 (1) (2022) 483.
[24] F. Roche, A. Somieski, S. Brandenburg, Behavioral changes to repeated takeovers in highly automated driving: effects of the takeover-request design and the nondriving-related task modality, Hum. Factors 61 (5) (2019) 839–849.
[25] S. Brandenburg, F. Roche, Behavioral changes to repeated takeovers in automated driving: the drivers' ability to transfer knowledge and the effects of takeover request process, Transport. Res. F Traffic Psychol. Behav. 73 (2020) 15–28.
[26] W. Payre, J. Cestac, P. Delhomme, Fully automated driving: impact of trust and practice on manual control recovery, Hum. Factors 58 (2) (2016) 229–241.
[27] C. Neubauer, G. Matthews, D. Saxby, Fatigue in the automated vehicle: do games and conversation distract or energize the driver? Proc. Hum. Factors Ergon. Soc. Annu. Meet. 58 (1) (2014) 2053–2057.
[28] Hongji Du, Xiaohua Zhao, Xingjian Zhang, Yunlong Zhang, Jian Rong, Effects of fatigue on driving performance under different roadway geometries: a simulator study, Traffic Inj. Prev. 16 (5) (2015) 468–473.
[29] S. Li, P. Blythe, W. Guo, A. Namdeo, Investigation of older driver's takeover performance in highly automated vehicles in adverse weather conditions, IET Intell. Transp. Syst. 12 (2018) 1157–1165.
[30] C. Gold, M. Körber, D. Lechner, K. Bengler, Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density, Hum. Factors 58 (4) (2016) 642–652.
[31] K. Zeeb, A. Buchner, M. Schrauf, What determines the take-over time? An integrated model approach of driver take-over after automated driving, Accid. Anal. Prev. 78 (2015) 212–221.
[32] N. Merat, A.H. Jamson, F.C.H. Lai, O. Carsten, Highly automated driving, secondary task performance, and driver state, Hum. Factors 54 (5) (2012) 762–771.
[33] S.H. Yoon, Y.G. Ji, Non-driving-related tasks, workload, and takeover performance in highly automated driving contexts, Transport. Res. F Traffic Psychol. Behav. 60 (2019) 620–631.
[34] F. Naujoks, C. Purucker, K. Wiedemann, C. Marberger, Noncritical state transitions during conditionally automated driving on German freeways: effects of non–driving related tasks on takeover time and takeover quality, Hum. Factors 61 (4) (2019) 596–613.
[35] F. Naujoks, D. Befelein, K. Wiedemann, A. Neukum, A review of non-driving-related tasks used in studies on automated driving. International Conference on Applied Human Factors and Ergonomics, 2017, pp. 525–537.
[36] S. Kim, R. van Egmond, R. Happee, Effects of user interfaces on take-over performance: a review of the empirical evidence, Information 12 (4) (2021) 162.
[37] M. Kuehn, T. Vogelpohl, M. Vollrath, Takeover times in highly automated driving, (level 3), in: 25th International Technical Conference on the Enhanced Safety of Vehicles, ESV) national highway traffic safety administration, 2017, pp. 1–11.
[38] S.H. Yoon, Y.W. Kim, Y.G. Ji, The effects of takeover request modalities on highly automated car control transitions, Accid. Anal. Prev. 123 (2019) 150–158.
[39] C. Gold, D. Damböck, L. Lorenz, K. Bengler, "Take over!" How long does it take to get the driver back into the loop? Proc. Hum. Factors Ergon. Soc. Annu. Meet. 57 (1) (2013) 1938–1942.
[40] A. Eriksson, N.A. Stanton, Takeover time in highly automated vehicles: noncritical transitions to and from manual control, Hum. Factors 59 (4) (2017) 689–705.
[41] A. Feldhütter, C. Gold, S. Schneider, K. Bengler, How the duration of automated driving influences take-over performance and gaze behavior, in: Advances in Ergonomic Design of Systems, Products and Processes, Springer, Berlin, Heidelberg, 2017.
[42] A. Bourrelly, C. Jacobé de Naurois, A. Zran, F. Rampillon, J.-L. Vercher, C. Bourdin, Long automated driving phase affects take-over performance, IET Intell. Transp. Syst. 13 (2019) 1249–1255.
[43] World Medical Association, World Medical Association Declaration of Helsinki: Ethical principles for medical tesearch involving human subjects, JAMA 310 (20) (2013) 2191–2194.
[44] J.F. Golding, Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness, Brain Res. Bull. 47 (5) (1998) 507–516.
[45] M.W. Johns, A New method for measuring daytime sleepiness: the Epworth Sleepiness Scale, Sleep 14 (6) (1991) 540–545.
[46] T.J. Smith, The attentional theory of cinematic continuity, Projections 6 (1) (2012) 1–27. Retrieved September 29th, 2022, from, https://www.berghahnjournals.com/view/journals/projections/6/1/proj060102.xml.

[47] E. Dogan, V. Honnet, S. Masfrand, A. Guillaume, Effects of non-driving-related tasks on takeover performance in different takeover situations in conditionally automated driving, Transport. Res. F Traffic Psychol. Behav. 62 (2019) 494–504.
[48] W.W. Wierwille, L.A. Ellsworth, Evaluation of driver drowsiness by trained raters, Accid. Anal. Prev. 26 (5) (1994) 571–581.
[49] S.M. Belz, G.S. Robinson, J.G. Casali, An on-road investigation of commercial motor vehicle operator self assessment of fatigue as an indicator of driver fatigue, Proc. Hum. Factors Ergon. Soc. Annu. Meet. 45 (23) (2001) 1576–1580.
[50] M.R. Yanko, T.M. Spalek, Driving with the wandering mind: the effect that mind-wandering has on driving performance, Hum. Factors 56 (2) (2014) 260–269.
[51] O. Jarosch, C. Gold, F. Naujoks, B. Wandtner, C. Marberger, G. Weidl, M. Schrauf, The Impact of Non-driving Related Tasks on Take-Over Performance in Conditionally Automated Driving – A Review of the Empirical Evidence. 9, Tagung Automatisiertes Fahren, 2019.
[52] N. Du, F. Zhou, E. Pulver, D.M. Tilbury, L.P. Robert, A.K. Pradhan, X.J. Yang, Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving, Transport. Res. C Emerg. Technol. 112 (2020) 78–87.
[53] J.A. Russell, A circumplex model of affect, J. Pers. Soc. Psychol. 39 (6) (1980) 1161–1178.
[54] S. Hergeth, L. Lorenz, J.F. Krems, Prior familiarization with takeover requests affects drivers' takeover performance and automation trust, Hum. Factors 59 (3) (2017) 457–470.
[55] M. Ebnali, K. Hulme, A. Ebnali-Heidari, A. Mazloumi, How does training effect users' attitudes and skills needed for highly automated driving? Transport. Res. F Traffic Psychol. Behav. 66 (2019) 184–195.
[56] S. Sibi, S. Balters, E. Fu, E.G. Strack, M. Steinert, W. Ju, Back to school: impact of training on driver behavior and state in autonomous vehicles, in: 2020 IEEE Intelligent Vehicles Symposium (IV), IEEE Press, 2020, pp. 1189–1196.
[57] S.E. Yoon, Y.W. Kim, Y.G. Ji, The effects of takeover request modalities on highly automated car control transitions, Accid. Anal. Prev. 123 (2019) 150–158.
[58] X. Tan, Y. Zhang, The effects of takeover request lead time on drivers' situation awareness for manually exiting from freeways: a web-based study on level 3 automated vehicles, Accid. Anal. Prev. 168 (2022) 106593.
[59] D. Davenne, R. Lericollais, P. Sagaspe, J. Taillard, A. Gauthier, S. Espié, P. Philip, Reliability of simulator driving tool for evaluation of sleepiness, fatigue and driving performance, Accid. Anal. Prev. 45 (2012) 677–682.
[60] J. Engström, E. Johansson, J. Östlund, Effects of visual and cognitive load in real and simulated motorway driving, Transport. Res. F Traffic Psychol. Behav. 8 (2) (2005) 97–120.
[61] M. Krause, L. Yilmaz, K. Bengler, Comparison of real and simulated driving for a static driving simulator, Adv. Hum. Aspects Transp.: Part II 8 (2014) 29.
[62] C. Bougard, D. Davenne, Effects of sleep deprivation and time-of-day on selected physical abilities in off-road motorcycle riders, Eur. J. Appl. Physiol. 112 (1) (2012) 59–67.