



HAL
open science

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

Arthur Portron, Gaëtan Perrotte, Guillaume Ollier, Clément Bougard,
Christophe Bourdin, Jean-Louis Vercher

► To cite this version:

Arthur Portron, Gaëtan Perrotte, Guillaume Ollier, Clément Bougard, Christophe Bourdin, et al..
Getting back in the loop: Does autonomous driving duration affect driver's takeover performance?.
Heliyon, 2024, pp.e24112. 10.1016/j.heliyon.2024.e24112 . hal-04411835

HAL Id: hal-04411835

<https://amu.hal.science/hal-04411835>

Submitted on 23 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

Journal Pre-proof

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

Arthur Portron, Gaëtan Perrotte, Guillaume Ollier, Clément Bougard, Christophe Bourdin, Jean-Louis Vercher

PII: S2405-8440(24)00143-9

DOI: <https://doi.org/10.1016/j.heliyon.2024.e24112>

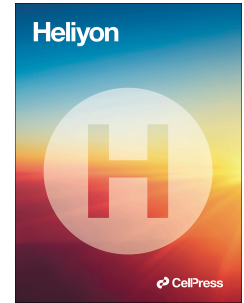
Reference: HLY 24112

To appear in: *HELIYON*

Received Date: 19 November 2023

Revised Date: 24 December 2023

Accepted Date: 3 January 2024



Please cite this article as: , Getting back in the loop: Does autonomous driving duration affect driver's takeover performance?, *HELIYON* (2024), doi: <https://doi.org/10.1016/j.heliyon.2024.e24112>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier Ltd.

1 **Getting Back In the Loop: Does autonomous driving duration affect driver's** 2 **takeover performance?**

3
4 Portron, Arthur^{a,*}; Perrotte, Gaëtan^{a,b}; Ollier, Guillaume^a; Bougard, Clément^b; Bourdin, Christophe^a and
5 Vercher, Jean-Louis^a

6
7 ^a Aix Marseille University, CNRS, ISM, Marseille, France

8 ^b Groupe Stellantis, Centre Technique de Vélizy, Vélizy-Villacoublay, France

9 * Corresponding author

10
11 **Keywords:** Takeover request; Driver behavior; Long durations; Non-critical event; Simulating level-3 automation
12

13 14 **Arthur PORTRON, Ph.D.**

15 Postdoctoral Researcher

16 UMR 7287 CNRS, Faculty of Sport Sciences

17 Aix-Marseille University, Marseille, France

18 Email : arthur.portron@gmail.com
19

20 **Gaëtan PERROTTE**

21 PhD candidate

22 UMR 7287 CNRS, Faculty of Sport Science

23 Aix-Marseille University, Marseille, France

24 Email: gaetan.perrotte@stellantis.com
25

26 **Guillaume OLLIER**

27 PhD Candidate

28 CEA Saclay, Gif-sur-Yvette, France

29 Email : guillaume.ollier@cea.fr
30

31 **Clément BOUGARD, Ph.D.**

32 Research Engineer

33 Stellantis Group

34 Poissy, France

35 Email : clement.bougard@stellantis.com
36

37 **Christophe BOURDIN, Ph.D.**

38 Full Professor

39 UMR 7287 CNRS, Faculty of Sport Sciences

40 Aix-Marseille University, 13009 Marseille, France

41 Email : christophe.bourdin@univ-amu.fr
42

43 **Jean-Louis VERCHER, Ph.D.**

44 Research Director at CNRS

45 UMR 7287 CNRS, Faculty of Sport Sciences

46 Aix-Marseille University, Marseille, France

47 Email : jean-louis.vercher@univ-amu.fr
48

49
50 **Abstract:** The level 3 autonomous driving function allows the driver to perform non-driving-related
51 tasks such as watching movies or reading while the system manages the driving task. However, when
52 a difficult situation arises, the driver is requested to return to the loop of control. This switching from
53 driver to passenger then back to driver may modify the driving paradigm, potentially causing an out-
54 of-the-loop state. We tested the hypothesis of a linear (progressive) impact of various autonomous
55 driving durations: the longer the level 3 autonomous function is used, the poorer the driver's takeover
56 performance. Fifty-two participants were divided into 4 groups, each group being assigned a specific
57 period of autonomous driving (5, 15, 45, or 60 minutes), followed by a takeover request with a time
58 budget of 8.3 s. Takeover performance was assessed over two successive drives via reaction times and
59 manual driving metrics (trajectories). The initial hypothesis (linearity) was not confirmed: there was a
60 nonlinear relationship between autonomous driving duration and takeover performance, with one
61 duration (15 minutes) appearing safer overall and mixed performance within groups. Repetition
62 induced a major change in performance during the second drive, indicating rapid adaptation to the
63 situation. The non-driving-related task appears critical in several respects (dynamics, content, driver
64 interest) to proper use of level 3 automation. All this supports previous research prompting reservations
65 about the prospect of car driving becoming like train travel.

66 1. Introduction

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

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

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

92 *1.1 Out-of-the-loop states*

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

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

112 Moreover, in manual mode, by activating the steering wheel and pedals, the driver is involved in a
113 physical loop of control of the vehicle, as well as in a cognitive loop of predicting, planning, and
114 deciding what to do in the driving environment [12]. During level 3 or higher phases of automation,
115 the driver switches voluntarily from the state of being “*in the loop*” to “*out of the loop*” (OOTL) through
116 a potential intermediate “*on the loop*” state [13], whereas this latter framework is never involved in
117 manual driving. It is now acknowledged that leaving the loops (of either cognitive or physical control
118 or both) is critical [14]. It can negatively impact the driver's performance when collaborating with

119 automated systems [15, 16, 17] and can result in a loss of SA, one of the deadliest forms of which is
120 drowsiness [18]. Furthermore, in a monotonous driving environment, the risk of fatigue and loss of SA
121 with automation is likely to be greater than under manual driving [19, 20]. To facilitate collaboration
122 between autonomous driving systems and human operators, it is crucial to identify factors inducing
123 the “*out-of-the-loop*” state in an autonomous driving environment as well as those which facilitate
124 getting “*back into the loop*”.

125 *1.2 Highly automated driving for long durations*

126 Several factors influencing the transition back to manual driving after a takeover request have been
127 extensively studied, particularly factors related to i) the human operator [21], like age [22,23],
128 experience [24,25,26], level of fatigue [27,28]), ii) the driving environment, like weather [29], traffic
129 [30], road type [28,31]), iii) non-driving-related tasks [32, 33, 34, see reviews 35, 36]. Studies have
130 also explored parameters related to the takeover request itself (e.g., modality of TOR [37, 38], time
131 budget (i.e. the time between TOR and impending collision) [39,40], and duration of autonomous
132 period before TOR [41, 42]).

133 However, for practical reasons, most of these studies were designed to expose the driver to short
134 periods of autonomous driving before a TOR, rarely involving their participants in really long
135 autonomous driving phases. Feldhutter et al. (2017) [41] showed that a period of 20 minutes of
136 automated driving increases response time to a takeover request but does not affect either acceleration
137 or time-to-collision profiles. Bourrelly et al. (2019) [42] tested even longer periods (60 minutes vs 10
138 minutes) and showed that overall takeover performance was poorer after the longer period, with a
139 longer reaction time and sharper trajectories during avoidance maneuvers. Vogelpohl et al. (2019) [20]
140 furthermore designed a protocol to compare automated driving versus manual driving under a
141 prolonged driving condition (60 min) or lack of sleep condition (< 5h sleep the night before the
142 experiment). They confirmed that in a given driving environment, autonomous driving shortens the
143 time before signs of fatigue appear. This effect was even more pronounced in some of the "lack of
144 sleep" drivers, who took longer to deactivate the autonomous function than the "long drive" group. To
145 summarize, though recent studies suggest that long autonomous driving phases could have deleterious

146 effects on driving performance, there is no consensus on which driving parameters are impacted. It is
147 therefore difficult to make strong recommendations to automotive manufacturers regarding whether
148 automation should be extended to longer periods or, on the contrary, limited to shorter periods. The
149 latter recommendation would run counter to the objectives behind highly automated functions.

150 *1.3 Aim of the study*

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

164

165 **2. Methods**

166 The present study was designed in order to provide cues about the points raised in the previous section.
167 Participants were required to drive in a monotonous environment. Then, rapidly they had to activate
168 the autonomous mode. Following this, they were informed they could quietly watch a movie on the
169 entertainment screen. They were also informed that at a given moment (without previously knowing
170 when) they would be requested to takeover, and they had to continue manually driving, whatever
171 happens. Though participants were previously trained to the takeover procedure (a single action on a

172 pedal or on the wheel will deactivate the automatic drive and give back control) they were never
173 exposed before the experimental session to the type of event used to test their performance. Thus, we
174 expected them to be surprised, and even more in case of inattention and eventually out-of-the loop
175 state. For this reason, we duplicated the experimental session (same duration, same event). Comparison
176 between the two sessions is expected to give cue about a sort of adaptation to the situation. In order to
177 limit this potential learning which would make comparison between durations difficult, we used
178 independent groups: each participant was exposed to only one duration, twice. To our knowledge this
179 is the first study testing more than two durations (actually four) including both short and long ones.
180 The protocol is detailed in the following sections.

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

184 In a static driving simulator located at the Virtual Reality Center facility of Aix-Marseille University,
185 51 participants were distributed in 4 independent groups (namely G05, G15, G45 and G60) in order to
186 reduce learning effect within the participants, each associated to different automated driving duration:
187 5 minutes, 15 minutes, 45 minutes or 60 minutes. They took part in conditional automated driving
188 scenarios divided into two recorded sequences (A and B) with a 10-minutes break between the two. In
189 each sequence, participants had to react to a takeover request to avoid a simulated road event in front
190 of the driver. During the autonomous driving period, participants were instructed to watch a movie and
191 were informed that the TOR, and therefore the road event, could arise at any time.

192 *2.1 Participants*

193 The participants were students at Aix-Marseille University, aged between 18 and 35 (Mean = 22.1; SD
194 = 3.66). Fifteen (33%) were females and 36 were males (66%). All had normal or corrected (by corneal
195 lenses) vision. Recruitment criteria were holding a valid driver's license for at least 6 months or driving
196 accompanied by a qualified driver for at least 2 years (only one participant), not wearing glasses, not
197 being susceptible to simulator sickness (assessed by the Motion Sickness Susceptibility Questionnaire
198 [44]), and having an Epworth scale score of 14 or less (assessing susceptibility to drowsiness [45]).

199 *Table 1. Summary of participants*

Groups	N	Age	Gender	Kms/year	Epworth
G05	13	22.4 (3.57)	F:4 ; M : 9	12800 (7777)	6.5 (3.26)
G15	13	22.7 (4.1)	F:3 ; M : 10	15000 (10100)	9.1 (2.50)
G45	13	21 (2.84)	F:4 ; M : 9	10962 (7885.8)	9.9 (3.84)
G60	12	22.6 (4.1)	F:4 ; M : 8	14.625 (7535.1)	8.75 (3.1)

200

201 *2.2 Group homogenization*

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

208 *2.3 Apparatus*209 *2.3.1 Simulator*

210 The static simulator consisted of a triple-screen visual setup (3 video screens, each 24" landscape
 211 format 16/9), a steering wheel from Stellantis Group[®], brake and accelerator pedals, and a small screen
 212 (10") used to display the dashboard, located just behind the wheel (see Figure1A). The driving
 213 environment was generated with SCANeR Studio Software[®] by AVSimulation[®], at a resolution of
 214 1920 x 1080 pixels onto each screen of the triple-screen, providing a 136° horizontal forward field of
 215 view. A rearview mirror was simulated in the central front screen, providing a 60° rear field of view,

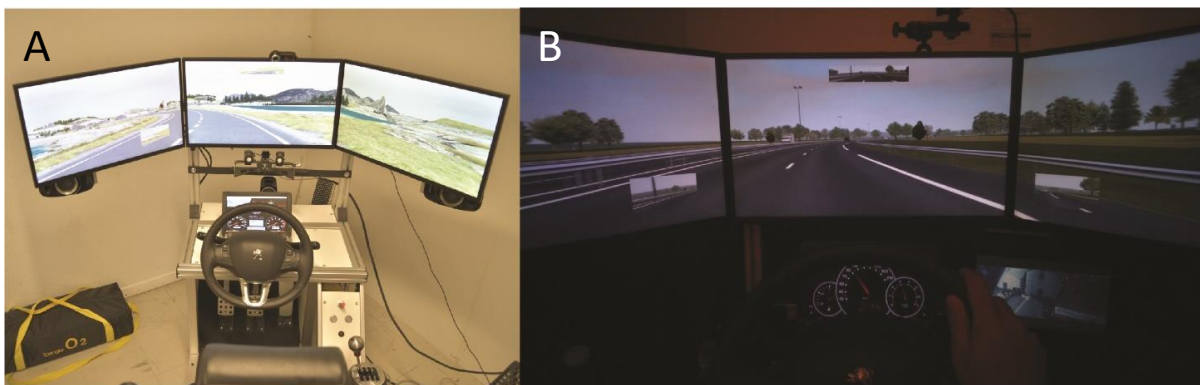


Figure 1. A. Top view of the static simulator. B. Perspective of a participant in the static simulator.

216 and two side mirrors were simulated respectively on the left and right screens. Two speakers were
217 placed under the left and right screens, while a subwoofer behind the simulator provided simulated
218 engine, road, and traffic sounds. A GoPro Hero7 camera located above the central screen (simulating
219 the driver's front view) recorded driver behavior in order to rate the driver drowsiness level in post hoc
220 analysis.

221 2.3.2 *Driving environment parameters*

222 For both recordings (sequences A and B), the road type was a 2x2 highway (See Fig. 1B). The
223 simulated landscape around the highway was relatively featureless, exposing participants to
224 monotonous driving conditions.

225 Traffic conditions were generated using the swarm generation function of SCANer Studio Software©
226 to create streaming traffic with the cars in randomly generated positions (range defined in parameters)
227 around the ego vehicle - defined as the queen of the swarm. Simulated weather conditions were a sunny
228 sky, and the experimental room had softened lights and a constant temperature ($20^{\circ} \pm 2$ Celsius).

229 Automation was provided by SCANer Studio Software© according to the current definition of a level
230 3 highly automated driving system [1]. When the autonomous function was activated, the system was
231 able to safely manage lateral trajectories and longitudinal and lateral speeds, to follow the course of
232 the road, and to keep a safe distance from vehicles ahead (time to collision > 2.50 seconds). The speed
233 during the level 3 automated period was set to 113 km/h, while in manual driving a warning message
234 appeared when the speed reached 120 km/h. The automated driving system was intentionally not
235 programmed to perform any overtaking maneuver. We want to make clear that we choose to conduct
236 the experiment without activating any ADAS brake system at the time of takeover, for two main
237 reasons: firstly, since we needed to identify appropriate/inappropriate manual maneuver after TOR and
238 take-over failure, and manual control trajectory were used as performance indicators, using ADAS
239 (e.g. automatic braking/collision avoidance) would have altered our capability to distinguish failure
240 from success. Secondly ADAS for automatic collision avoidance could induce a bias in the driver's
241 behavioral response.

242

243 2.3.3 *Non-driving-related task*

244 To ensure that participants did not anticipate the event and were involved in fairly uniform behavior at
245 the time of the TOR, and to heighten their OOTL state, all participants were involved in a non-driving
246 secondary task: watching a movie. This was intended to keep them busy during the session, since doing
247 nothing for up to one hour is unrealistic. They had to watch a movie (a blockbuster called "Aquaman";
248 2018, WarnerBros ©) aimed at attracting and holding viewers' attention [46].

249 2.3.4 *Experimental procedure*

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

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

261 2.3.4.1 *Familiarization phase*

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

265 In a semi-urban environment, the participants had to follow a lead car during a manual driving period
266 of approximately six minutes. Then the system sent a Manual-to-Autonomous (MtA) notification

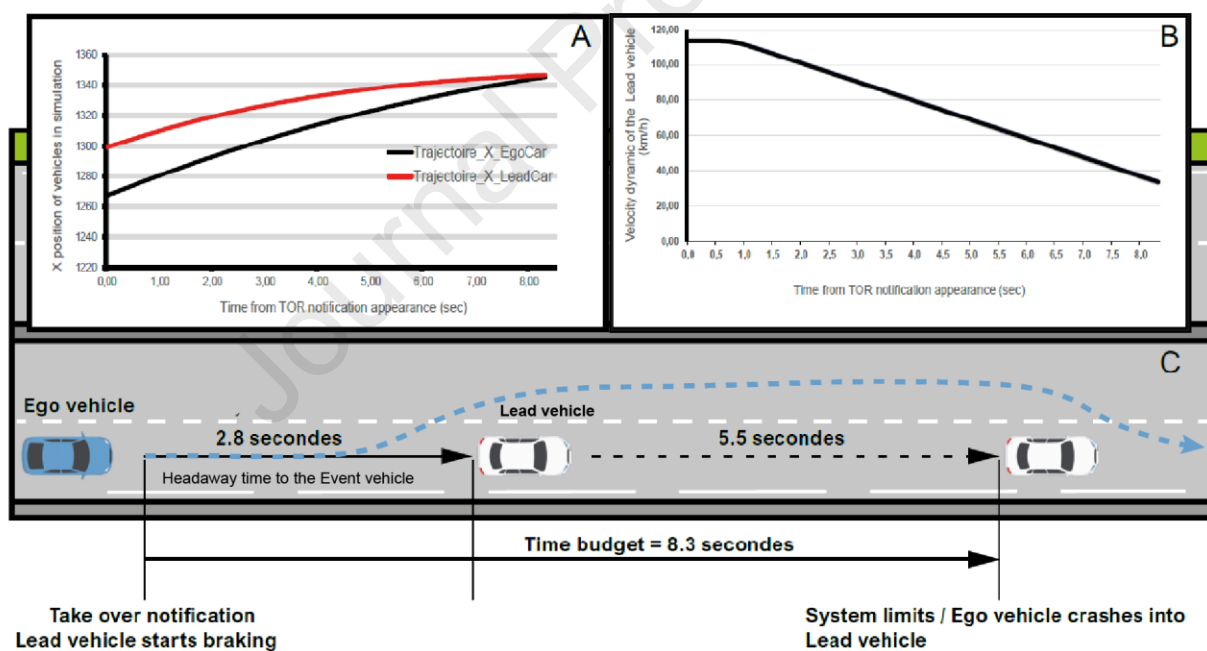
267 requiring the participants to activate the autonomous function. Thirty seconds after this MtA driving
268 transition, the system displayed the first TOR involving a critical event with a takeover time (TOT) of
269 4.95 seconds. This was the sudden appearance of a police car stopped in the right lane and a car in the
270 security lane, with two surrounding traffic cones. After avoiding this first event (no crashes recorded
271 among the participants), participants had to drive manually for ninety seconds. After a second
272 autonomous period of about three minutes and thirty seconds, there was another TOR involving a
273 critical event (TOT = 3 seconds): a pick-up car located in the middle of a crossroads (no crashes
274 recorded for any participants). Then participants had again to drive manually for one more minute,
275 thus reaching the end of the familiarization phase. Both time budgets were critical, aimed at inducing
276 a high level of expectation about the criticality and uncertainty of TOR events in the subsequent
277 sequences.

278 *2.3.4.2 Main recording sequences*

279 Since the two experimental sequences (A and B) were performed under the same conditions, we only
280 describe here one of them. The simulated environment was a highway ring. The drivers were in a
281 stationary position in the middle of the traffic and they had to reach the speed limit while keeping a
282 safe distance from the lead car (2 security-lanes), overtaking being forbidden. After a period of manual
283 driving (2 minutes), the system sent a MtA request. After turning on the autonomous driving function,
284 the participants were instructed to watch a movie on the screen located to the right of the steering
285 wheel. They were allowed only to manage the sound level; no timestamp was displayed, to avoid
286 participants counting the time and to prevent any anticipation. At the end of the autonomous period
287 (lasting 5, 15, 45 or 60 minutes depending on the group), the system sent a TOR notification. The
288 critical event involved here differed somewhat from those used in the familiarization phase. For the
289 experimental sequences, the front car suddenly started braking from 113 km/h to 33 km/h in 5.5
290 seconds, providing a time budget of 8.3 seconds before collision at constant speed (see Figure 2A and
291 2B ; all traffic other than the event vehicle was removed). This time budget is recommended by several
292 authors [38,39] as sufficient to ensure safe return to manual control of the car. A participant who did
293 not detect the TOR notification would collide with the event vehicle after the 8.3 seconds, ending the

294 session. Thus we consider that a successful trial means the driver was able to avoid collision while
 295 staying on the road, by either braking and changing lane, or just avoiding the obstacle by changing
 296 lane (see Figure 2C), no specific instruction about what to do was given. A participant successfully
 297 deploying an avoidance maneuver would drive manually for 1 more minute.

298 After the first sequence (A), participants had a 10-minute break, left the simulator to stretch their legs,
 299 then went back for the second sequence (B). The only parameter differing from sequence A was the
 300 starting point of the movie (participants watched the movie from the point where they had stopped it
 301 in sequence A). The only thing that the participant knew about the task in both experimental sequences
 302 was that at some point there would be a TOR notification involving a driving event that might be
 303 avoided, before driving on manually. These instructions purposely created ambiguity and uncertainty
 304 for the participant, aiming to simulate an "ecological" use of the level 3 automated driving function.



305 *Figure 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.*

306

307 2.3.5 Takeover performance assessment

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

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

317 *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 meter 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 0.5m 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 0.5m 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.

318

319 2.3.6 Assessment of drowsiness over time

320 The level of drowsiness was determined according to a method proposed by Wierville and Ellsworth
 321 [48] and based on subjective assessment by video analysis. We used the Observer Rating of Drowsiness
 322 (ORD) scale [49], with levels of drowsiness numbered (level 0: alert state, to level 4: extremely
 323 drowsy) for each minute of the driving sequences and for each subject recorded. Due to the general
 324 reliability of paired ratings by the same rater, there was only one rating per recording.

325 2.3.7 *Statistical Analysis*

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

332

333 **3. Results**

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

337 3.1 *Crash rates by autonomous driving duration*

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

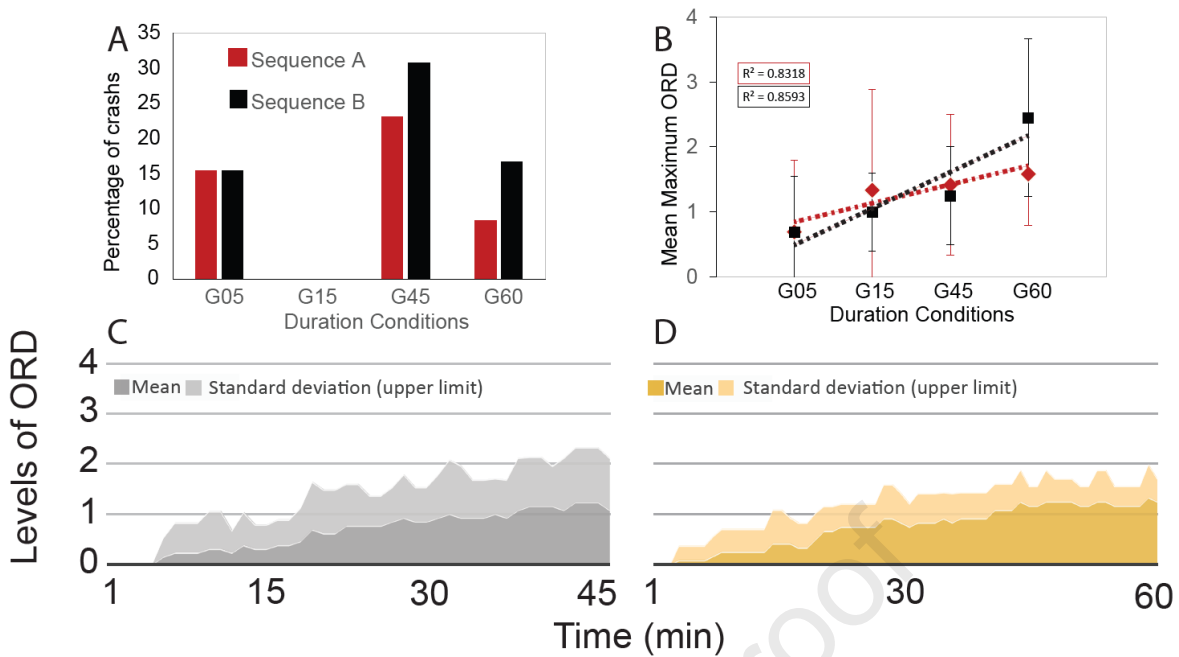


Figure 3. Percentage of crashes observed for each group and each sequence (Panel A). Mean maximum ORD rated for each group in sequence A (red) and sequence B (black, Panel B). Bottom panels illustrate mean assessment of ORD, for group G45 (Panel C) and G60 (Panel D). Dark areas in both panels illustrate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

346

347 3.2 Drowsiness-level ratings over time

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

355 3.3 Time to takeover

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

358 *3.3.1 Impact of duration of autonomous driving*359 *3.3.1.1 First drive (Sequence A)*

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

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

370

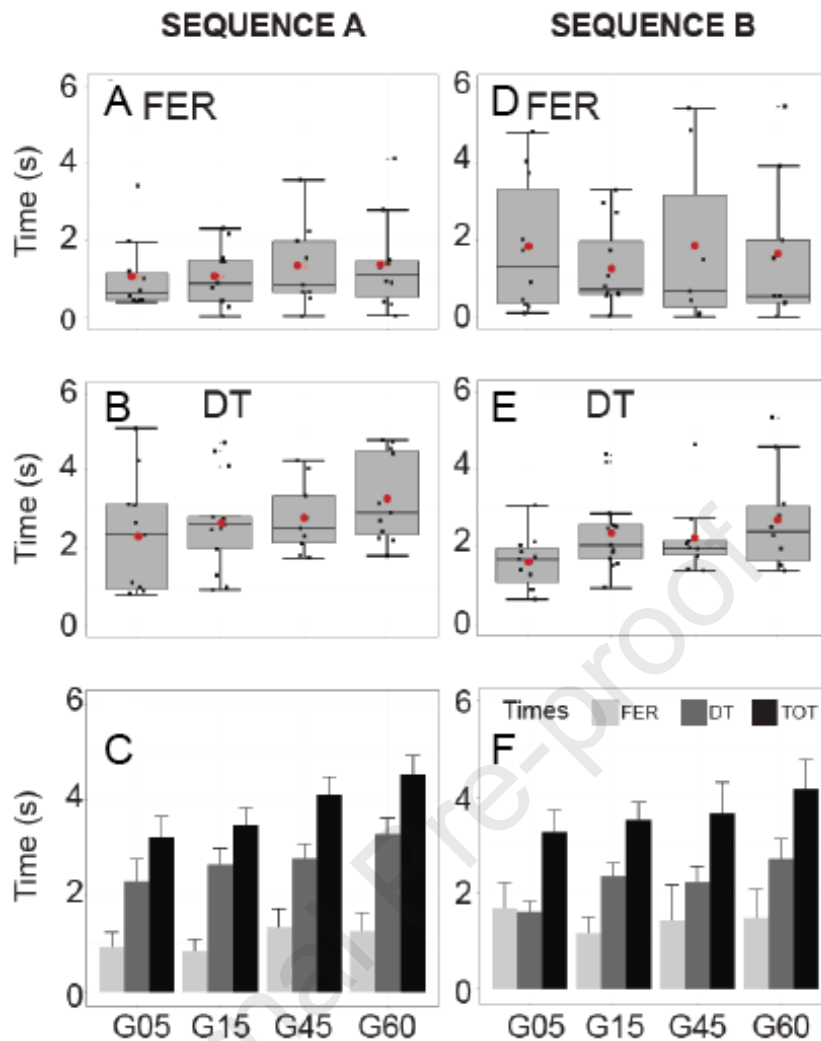


Figure 4. Reaction times measured for sequence A (left panels, A, B, C) and sequence B (right panels, D, E, F). Panels A, D: boxplot illustrating the distribution of first eye on road time (FER). Panels B, E: boxplot illustrating decision time (DT) showing distribution of measures across groups. Colored 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.

371

372

3.3.1.2 Second drive (Sequence B)

373 Like sequence A, the three Kruskal-Wallis tests conducted did not reveal any statistical difference

374 between groups (see Table 3, all p-values > 0.05) and visual inspection of distribution of FER times

375 did not reveal any effect of duration (Fig. 4D). Unlike sequence A, however, decision time (Fig.4E)

376 did not follow a linear trend. Decision times were shorter for G05, but similar for G15 and G45, and

377 longer for G60. A positive linear trend was observed (as with sequence A) for mean takeover times

378 (Fig. 4F; from 3.22 ± 1.54 s for G05 to 4.18 ± 2.15 s for G60, see Table 3).

379 *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	<i>p</i>
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

380

381 *3.3.1.3 Comparison between sequence A and sequence B*

382 Comparisons between sequences A and B for each group showed no difference (Wilcoxon test: $p >$
383 0.05): the groups' times remained fairly stable from sequence A to sequence B (see Fig.4C and F).

384 *3.3.2 Pooled data*

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

389 *3.3.2.1 Sequence A*

390 Comparison between data from Under 30 and Over 30 (see Fig. 5A) showed a significant difference
391 in takeover time between the groups (Fig. 5, left panel). Analysis revealed significantly longer takeover
392 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, $p =$
393 0.019), with a moderate effect size ($r = 0.351$).

394 *3.3.2.2 Sequence B*

395 In sequence B, visual inspection (Fig. 5B) showed a similar pattern, with quite similar FER and slightly
396 greater DT and TOT for the Over 30 minutes group. Statistical analysis did not reveal any significant
397 difference in the three reaction times between Under 30 minutes and Over 30 minutes. Subsequent
398 analysis revealed that the absence of statistical difference (compared to sequence A) for DT is due to
399 the high standard deviation measured for the >30 and the more similar means of each group ($TOT_{<30}$
400 $= 3.41 \pm 1.40$ s; $TOT_{>30} = 3.94 \pm 1.90$ s).

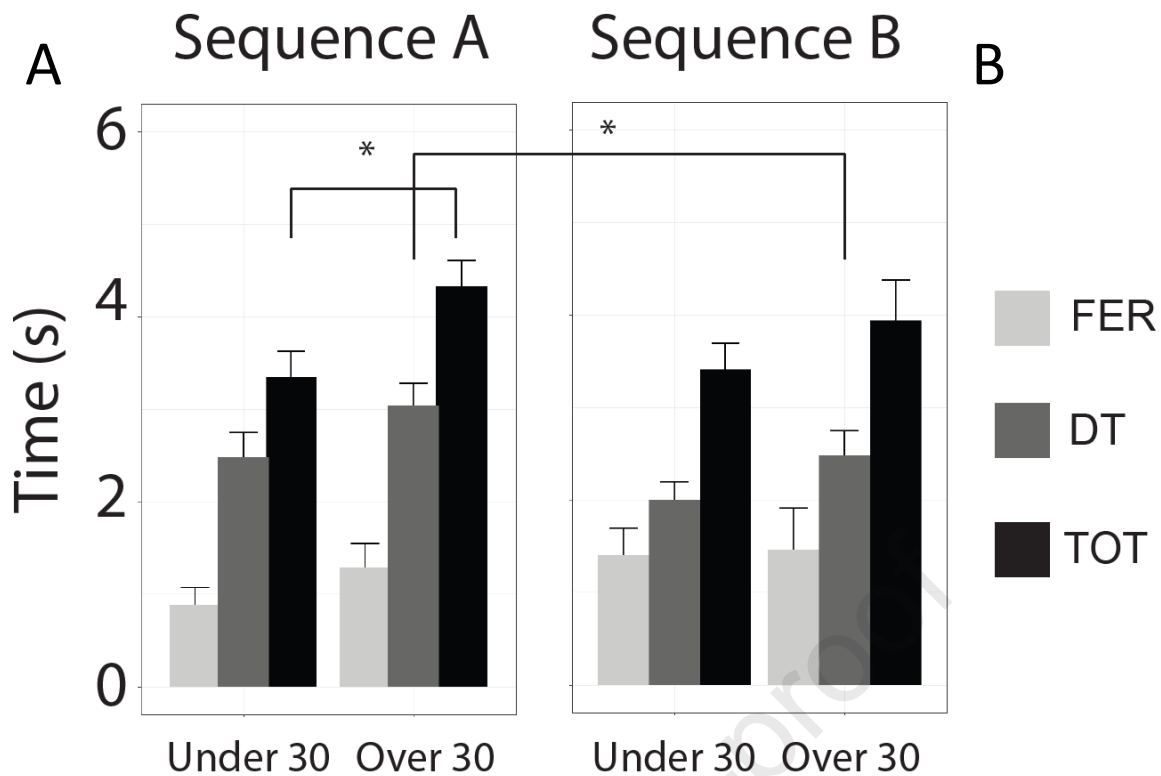


Figure 5. All times for groups Under 30 minutes and Over 30 minutes in both sequences, Panel A for sequence A, Panel B for sequence B. Bar plots show mean values for each time and each group. Significant differences are indicated by a star ($p < 0.05$). Abbreviations are defined on Table 2.

401

402

3.3.2.3 Comparison between sequence A and sequence B

403 An inter-sequence pairwise comparison on times for each group revealed that DT was significantly
 404 shorter in sequence B than in sequence A for the Over 30 minutes group ($Ov30_{dtA} = 3.04 \pm 1.05s$);
 405 $Ov30_{dtB} = 2.48 \pm 1.17s$); Wilcoxon test: $p < 0.05$).

406

3.4 Avoidance manoeuvre and quality of control

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

409

3.4.1 Influence of autonomous driving duration

410 As shown in the panels A,C,E and G of Fig. 6, lateral deviations to manage the situation varied widely
 411 in amplitude in all groups for sequence A. Inter-individual variability was mostly due to variable RTs,
 412 also some drivers braked before changing lane, others did not (see speed variability on Fig. 6B, 6D,
 413 6F and 6H respectively for group G05, G15, G45 and G60). Consequently, deviations occurred at

414 different times and some participants produced more than one deviation. Thus there are several peaks
 415 in the average curves, the highest at ten seconds means this is the most probable time when an initial
 416 deviation occurred. A successful reaction (no crash) means necessarily the driver changes lane. Indeed,
 417 as stated in the methods section, each lane width is 3.5 m. Thus a lateral deviation of at least 3 m means
 418 a lane change (occurring in average 10 s after takeover).

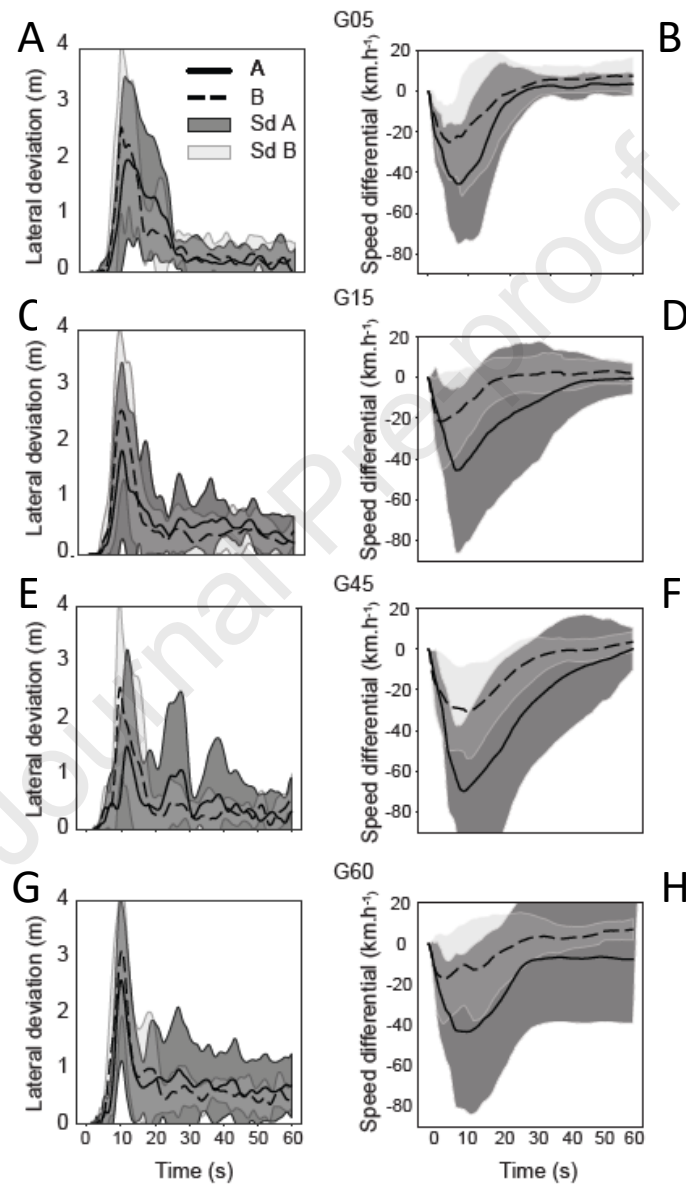
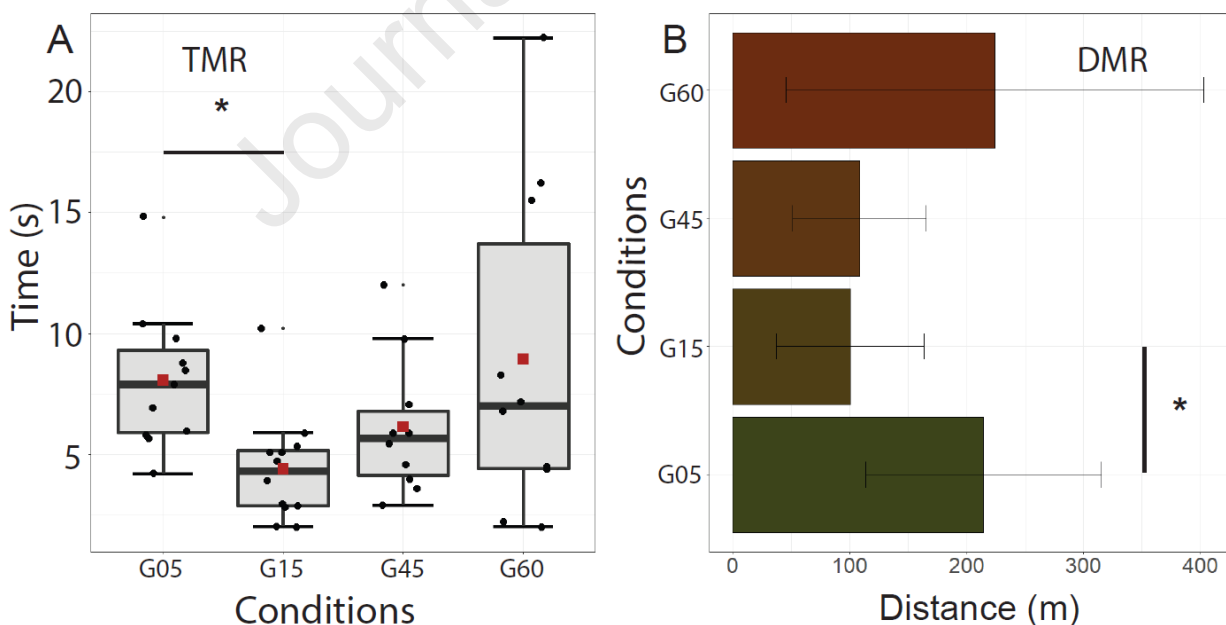


Figure 6. Average lateral deviation from initial lateral position of the car at the start of manual driving (left panels: A, C, E, and G) and average speed differential from start of manual driving (right panels : B, D, F and H) for each group (from top to bottom : G05, G15; G45 and G60). Deactivation 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).

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

426 On no indicator (see 2.3.5) did distributions respect the normality assumption (Shapiro-Wilk test, p
 427 <0.05), and therefore Kruskal-Wallis tests were used for comparisons between groups in sequence A
 428 and sequence B. Of the five indicators (see Table 4), only two (TMR and DMR) showed a statistically
 429 significant difference relative to duration (see Fig. 7), in sequence A only (none in sequence B). The
 430 time taken to return to the initial position in the right lane after reaching the maximum deviation (TMR)
 431 was influenced by the duration of the autonomous drive (Kruskal-Wallis test; $p = 0.02$; $\eta^2 = 0.159$
 432 (large effect size)). Post-hoc Kruskal-Nemenyi tests indicated a significantly longer time for group
 433 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).



434
 435 *Figure 7. Illustration of time (Panel A, TMR = Time from Maximum deviation to Right lane position) and*
 436 *distance (Panel B, DMR = Distance from Maximum deviation to Right lane position) to return to initial*
 437 *right lane position after reaching maximum lateral deviation during sequence A. Panel A: colored squares*
 438 *show the mean for each group in each boxplot, black dots show individual measurements. Horizontal bars*

439 at the ends represent standard deviation, central black horizontal lines represent median for each group.
 440 Panel B: Horizontal black lines for each bar plot illustrate standard deviation. Significant differences are
 441 indicated by a star ($p < 0.05$).

442 In line with the previous results, a significant difference between groups was observed in sequence A
 443 (Kruskal-Wallis test; $p = 0.01$; $\eta^2 = 0.184$ (large effect size)) on the distance covered by drivers from
 444 the maximum deviation position until they returned to the initial right lane position (DMR). Post-hoc
 445 Kruskal-Nemenyi tests revealed ($p = 0.022$) a significantly longer distance for group G05 ($M_{DMR} =$
 446 214.35 ± 100.55 m) than for group G15 ($M_{DMR} = 100.53 \pm 63.40$ m) (Fig.7B). The last three indicators,
 447 which did not show any difference between the initial four autonomous driving duration groups, were
 448 merged, as previously for times, into two groups: Under 30 minutes and Over 30 minutes. When data
 449 were compared between these two groups, no statistical difference was observed (Wilcoxon test;
 450 $p > 0.05$). The following section thus presents inter-sequence comparisons with all data pooled.

451 Table 4. Ego car measurements taken into account to assess takeover performance during maneuver. Each cell pools
 452 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	0.290 (3.88)	-0.258 (7.77)	-3.389 (8.46)	1.09 (5.94)	3.01	0.39
SEQUENCE B	2.54 (4.08)	3.62 (1.86)	2.55 (2.50)	4.09 (3.71)	4.04	0.25
MD ($m \cdot s^{-2}$)						
SEQUENCE A	-9.36 (3.51)	-6.26 (4.64)	-9.13 (4.10)	-6.08 (4.46)	6.12	0.1
SEQUENCE B	-7.01 (4.73)	-5.23 (4.69)	-5.67 (4.73)	-4.02 (3.67)	3.38	0.33
TMD (s)						
SEQUENCE A	9.92 (3.94)	10.26 (7.91)	13.88 (9.18)	9.50 (7.19)	3.14	0.36
SEQUENCE B	7.69 (4.27)	6.68 (2.07)	7.32 (2.88)	6.53(3.97)	1.71	0.63
TMR (s)						
SEQUENCE A	8.07 (2.93)	4.40 (2.26)	6.12 (2.85)	8.93 (6.77)	9.37	0.02
SEQUENCE B	6.23 (3.44)	6.89 (7.77)	5.61 (3.33)	7.25 (5.02)	1.90	0.59
DMR (m)						
SEQUENCE A	214.35 (100.55)	100.53 (63.40)	108.37 (57.30)	223.92 (178.76)	10.35	0.01
SEQUENCE B	176.77 (107.17)	202.81 (253.24)	199.62 (154.78)	122.54 (65.95)	2.052	0.56

453 3.4.2 Comparison between Sequences A and B

454 Analyses of lateral deviation and speed differential profiles (see Figure 6) showed that, in all groups,
 455 behaviors at takeover differed between sequences A and B. In sequence A, some participants perceived
 456 the TOR earlier, quickly braked to deactivate the function, but took some time before starting the
 457 avoidance maneuver and changing lanes. Others braked sharply on detecting the TOR, then waited
 458 briefly before overtaking the lead car. In other words, there was a wide range of takeover behaviors in
 459 all our groups in sequence A. However, behaviors differed less in sequence B, with a smaller standard

460 deviation in all groups. In the light of the major behavioral discrepancies between sequence A and
 461 sequence B, we conducted an inter-sequence comparison (Fig. 8).

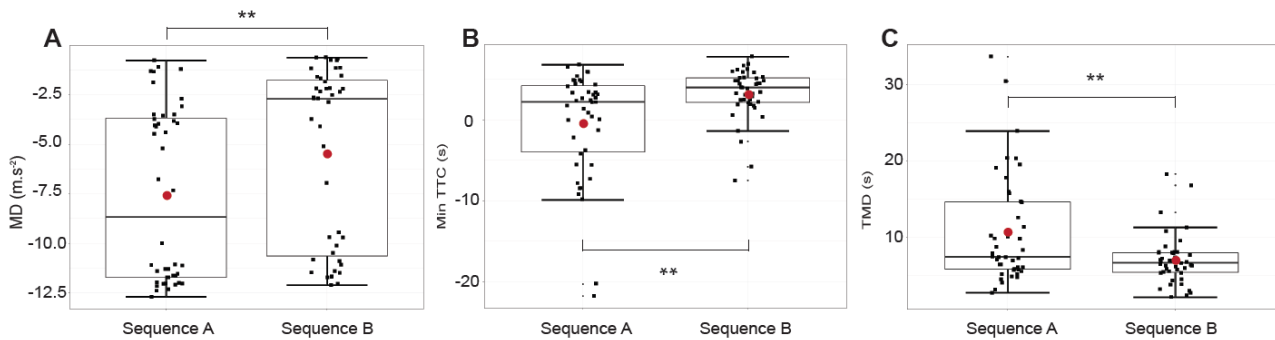


Figure 8. Comparison of Maximum deceleration (Panel A), Minimum TTC (Panel B), and Time to reach maximum deviation (Panel C) between sequence A and sequence B. Red dots show mean value for each measure. Significant differences are indicated by double stars ($p < 0.01$). Abbreviations are defined on table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

462 3.4.2.1 Maximum deceleration (MD)

463 The distribution of the values (Fig.8A) revealed two different performance profiles in both sequences,
 464 with one set of participants presenting deceleration values above -5 m.s^{-2} and another presenting values
 465 below -5 m.s^{-2} . A Mann-Whitney test conducted on maximum deceleration revealed a significant
 466 difference between sequences A and B ($MD_{SA} = -7.58 \pm 4.36 \text{ m/s}^2$; $MD_{SB} = -5.40 \pm 4.46 \text{ m/s}^2$; $p < 0.01$),
 467 with a moderate effect size ($r = 0.320$). This difference indicates that, in addition to showing two
 468 different speed profiles during the maneuver, drivers tended to slow down less during the maneuver in
 469 sequence B than in sequence A.

470 3.4.2.2 Minimum time to collision (Min TTC)

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

476 *3.4.2.3 Time to reach maximum deviation (TMD)*

477 As with the previous variables, the time taken to reach maximum deviation differed significantly
478 between sequences A and B (Fig. 8C, $TMD_{SA}=11.24\pm7.04$ s; $TMD_{SB}=6.99\pm3.36$ s); $p < 0.01$, with a
479 large effect size ($r = 0.763$). Thus drivers did their avoidance maneuver faster in sequence B than in
480 sequence A.

481 **4. Discussion**

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

490 *4.1 Impact of autonomous driving duration on takeover performance*

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

500 Strictly speaking, the assumption of linearity should have led to the observation of a greater number
501 of collisions with longer autonomous phases, with a maximum ORD at the time of TOR for the longer

502 duration groups. However, the crash rate was not correlated with the mean maximum ORD. Even
503 though the maximum ORD was often reported in the longer duration groups, it did not appear
504 systematically at the critical moment of TOR. The crash rate was highest in G45, followed by G60,
505 but there were also some crashes in the shorter duration groups in both sequences (see Fig. 4A). All
506 these observations are in contradiction with the supposed linear influence of duration of autonomous
507 driving on takeover performance and safety. This highlights the weight and the role of the NDRT and
508 individual factors, like fatigue due to lack of sleep, which will be discussed in the following sections.

509 *4.1.1 Are reaction times an indicator of performance?*

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

515 Results from pooled data show that the mean takeover time (from TOR to action) was impacted by the
516 duration of the autonomous drive. Exposure to more than 30 minutes in level 3 autonomous driving
517 increased the time taken to deactivate the function and regain in-the-loop control (manual driving).
518 This confirms and reinforces previous findings by Bourelly et al., [42] on the effect of prolonged
519 autonomous driving on time taken to react to a TOR. They reported an increase of 0.5 s in takeover
520 time for a 60-min duration relative to a 10-min period (paired groups), while we measured a mean
521 difference greater than 1 s between takeover times for durations of under 30 minutes and those for
522 durations of over 30 minutes. According to the literature on factors influencing times of action in
523 driving tasks, an increase of 1 second is of critical concern from an ecological perspective. The absence
524 of effect of autonomous driving duration on times in sequence B using the pooled data may be
525 explained by the change in decision time between the two sequences for the over 30 minutes group.
526 Decision time became quicker in sequence B and closer to that measured for the under 30 minutes
527 group in sequence B, likely due to better decision-making when faced with the event. Thus, while the
528 times measured here suggest some influence from the duration of autonomous driving, they are not

529 necessarily informative about performance. These times should be considered in the light of the time
530 budget of the takeover situation. Autonomous driving duration likely interacts with other factors
531 inherent to the driver or the environment to determine performance.

532 4.1.2 *Indicators of manual driving performance*

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

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

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

561 *4.2 Influence of the non-driving-related task*

562 The NDRT we used may have influenced takeover performance: a superhero action movie,
563 “Aquaman” (2018, WarnerBros ©). It was chosen because it was long enough to be watched over the
564 two driving sequences (especially given groups G45 and G60). This blockbuster follows a specific
565 format alternating action phases and calmer periods. One of the most important aspects of an NDRT
566 is what Jarosch et al. [51] defined as the way it affects the driver’s arousal level, and thus how it affects
567 cognitive transition (the return to being “On the loop” of control). This is especially important when
568 the participant’s attention is fully devoted to the task, as with the movie in the present study. Du et al.
569 [52] demonstrated that performance at takeover is correlated to the emotional state of the driver,
570 reporting that positive emotions induce a shorter takeover time, whereas level of arousal has no
571 influence. This leads us to suggest that watching this type of movie could, depending on the scene the
572 driver is watching at TOR, trigger either a positive or a negative emotion (as described by Russel [53])
573 and thereby modulate the individual’s takeover performance. Indeed, the scenes our participants were
574 watching just prior to the TOR differed substantially between groups and sequences. The content
575 varied in terms of temporality, dynamics, and subjects, with slow scenes, suspenseful scenes, several
576 scenes consisting of dialog, and fight scenes. In G15 (no crashes), the scenes watched during the last
577 minute before the TOR were fight scenes containing several high intensity events, in both sequences.
578 In G45’s sequence A (crash rate of 23%), the scene watched just before the TOR was a dialog scene
579 with no action but revealing an important plot event, whereas in sequence B (crash rate of 31%), it was
580 a fight scene. The same pattern was observed for G60: a dialog scene in sequence A and a fight scene

581 in sequence B, with more accidents in B than in A. Moreover, G05, where no fight scene was watched
582 just prior to TOR, but a storytelling scene in sequence A and a suspenseful scene in sequence B, showed
583 a consistent crash rate (15%) between sequences. Such differences in the type of scenes being watched,
584 with the unknown emotional valence induced by these scenes, may add to the idiosyncrasies in
585 takeover performance. Further, the task “watching a movie” could have been approached differently
586 by our participants, depending on their personal interest in this kind of movie and their perception of
587 movie-watching itself: for some, it may have been a monotonous or boring task, whereas for others it
588 was an exciting task.

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

594 *4.3 Effect of repetition*

595 The procedure used here revealed that repetition reduced variability, likely due to a process of
596 adaptation or learning by participants. As reported by previous studies showing an improved return to
597 the manual control loop [47], we found a clear difference in performance in sequence B for participants
598 who had already successfully resumed control of the vehicle, as well as a better understanding of the
599 situation. These observations are supported by the difference in control resumption behaviors between
600 sequence A and sequence B. In sequence A, some participants detected the TOR, resumed control of
601 the vehicle by strongly braking (high maximum deceleration), then waited a little before executing the
602 overtaking maneuver (longer decision time and minimum time-to-collision exceeding the time budget).
603 In sequence B, velocity profiles showed that participants resumed control of the vehicle by lightly
604 braking and immediately changing lane at high speed. This evolution is in line with Gold et al. [38],
605 who reported that the longer drivers had to decide what to do, the less they used their brakes (see Fig.
606 6D). This pattern of behavior suggests a potential adaptation and transfer of what they learned from

607 prior exposure [54,55] to the same situation, which is also consistent with the difference observed in
608 decision time between sequence A and B.

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

613 *4.4 Limitations*

614 Certain limitations should be taken into account when interpreting the present results. The first is the
615 level of expertise and the age of our participants with regard to the autonomy level studied here, likely
616 to be implemented in cars purchased by older people. Being less experienced in using high-level
617 driving aids or new assisted driving technologies could increase variability in behavioral responses.
618 Second, we identified two issues concerning the non-driving-related task. We did not control the
619 emotional valence for individual participants of the movie scenes watched just prior to takeover.
620 Moreover, watching a film during a long trip might not be a familiar task to all subjects, underlining
621 their motivational heterogeneity.

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

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

633 comparable way [59]. In addition, it is also well known that simulator studies emphasize driving
634 impairments noticeably regarding attention and drowsiness consequences.

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

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

642 **5. Conclusion**

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

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

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

657

658 **Author contributions**

- 659 • **Arthur Portron:** Methodology, Software, Formal analysis, Investigation, Resources, Data
660 Curation, Writing Original Draft, Visualization
- 661 • **Gaetan Perrotte:** Formal Analysis, Writing-Review & Editing,
- 662 • **Guillaume Ollier:** Software, Data Curation
- 663 • **Clément Bougard:** Conceptualization, Writing-Review & Editing
- 664 • **Christophe Bourdin:** Conceptualization, Resources, Writing-Review & Editing
- 665 • **Jean-Louis Vercher:** Conceptualization, Methodology, Investigation, Resources, Writing-
666 Review & Editing, Supervision, Project administration, Funding acquisition.

667 **Declaration of competing interest**

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

670 **Acknowledgment**

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

675

676 **References**

- 677 [1] Committee, S.O.R., & others, 2021. Taxonomy and definitions for terms related to driving
678 automation systems for on-road motor vehicles. SAE International: Warrendale, PA, USA.
- 679 [2] Mercedes Benz, 2022. The front runner in automated driving and safety technologies.
680 <https://group.mercedes-benz.com/innovation/case/autonomous/drive-pilot-2.html>
- 681 [3] Honda, March 4, 2021. Honda to begin sales of Legend with new Honda SENSING Elite.
682 <https://global.honda/newsroom/news/2021/4210304eng-legend.html>
- 683 [4] Audi A. G., 2016. Audi piloted driving. [https://www.audi-mediacycenter.com/en/audi-at-the-ces-](https://www.audi-mediacycenter.com/en/audi-at-the-ces-2016-5294/piloted-driving-5303)
684 [2016-5294/piloted-driving-5303](https://www.audi-mediacycenter.com/en/audi-at-the-ces-2016-5294/piloted-driving-5303)
- 685 [5] Volvo Cars, 2016. Intellisafe Autopilot. Intellisafe Autopilot.
- 686 [6] Bradley, R., 2016. Tesla Autopilot. MIT Technology Review.
687 [https://www.technologyreview.com/s/600772/10-breakthrough-technologies-2016-tesla-](https://www.technologyreview.com/s/600772/10-breakthrough-technologies-2016-tesla-autopilot/)
688 [autopilot/](https://www.technologyreview.com/s/600772/10-breakthrough-technologies-2016-tesla-autopilot/)
- 689 [7] Carsten, O., Martens, M.H., 2019. How can humans understand their automated cars? HMI
690 principles, problems and solutions. Cogn. Tech. Work. 21, 3–20.
- 691 [8] Cameron, J.A., Sottile, A.J., 1997. Automation effects on a driver's vigilance in the automated
692 highway system, Volume 1. Final report. Tech. rep., Crew System Ergonomics Information
693 Analysis Center Wright-Patterson AFB OH.
- 694 [9] Merat, N. Jamson, A., 2009. How do drivers behave in a highly automated car?, 5th Driving
695 Assessment Conference, pp. 514-521.
- 696 [10] Sarter, N.B., Woods, D.D., 1991. Situation awareness: a critical but ill-defined phenomenon, The
697 International Journal of Aviation Psychology. 1 (1), 45-57.
- 698 [11] Endsley, M.R., Bolstad, C.A., 1994. Individual differences in pilot situation awareness, The
699 International Journal of Aviation Psychology. 4 (3), 241-264.

- 700 [12] Boer, E.R., Hoedemaeker, M., 1998. Modeling driver behavior with different degrees of
701 automation: A hierarchical decision framework of interacting mental models. Proceedings of
702 17th European annual conference on human decision making and manual control, pp. 63–72.
- 703 [13] Merat, N., Seppelt, B., Louw, T. et al., 2019. The “Out-of-the-Loop” concept in automated driving:
704 proposed definition, measures and implications. *Cogn. Tech. Work.* 21, 87–98.
- 705 [14] Endsley, M.R., 2006. Situation awareness. In *Handbook of Human Factors and Ergonomics*, G.
706 Salvendy (Ed.). John Wiley & Sons.
- 707 [15] Nilsson, J., P. Falcone, P., Vinter, J., 2015. Safe transitions from automated to manual driving
708 using driver controllability estimation. *IEEE Transactions on Intelligent Transportation*
709 *Systems.* 16 (4), 1806-1816.
- 710 [16] Kaber, D.B., Endsley, M.R., 1997. Out-of-the-loop performance problems and the use of
711 intermediate levels of automation for improved control system functioning and safety. *Proc.*
712 *Safety Prog.* 16, 126-131.
- 713 [17] Berberian, B., Somon, B., Sahai, A., Gouraud, J., 2017. The out-of-the-loop bBrain: a
714 neuroergonomic approach of the human automation interaction. *Annual Reviews in Control.*
715 44, 303–315.
- 716 [18] Horne, J., Reyner, L., 1999. Vehicle accidents related to sleep : A review. *Occupational and*
717 *Environmental Medicine.* 56 (5), 289.
- 718 [19] Schömig, N., Hargutt, V., Neukum, A., Petermann-Stock, I., Othersen, I., 2015. The interaction
719 between highly automated driving and the development of drowsiness. *Procedia*
720 *Manufacturing.* 3, 6652–6659.
- 721 [20] Vogelpohl, T., Kühn, M., Hummel, T., Vollrath, M., 2019. Asleep at the automated wheel.
722 Sleepiness and fatigue during highly automated driving. *Accident Analysis & Prevention.* 126,
723 70–84.

- 724 [21] Merat, N., de Waard, D., 2014. Human factors implications of vehicle automation: Current
725 understanding and future directions. *Transportation Research. Part F: Traffic Psychology and*
726 *Behaviour*. 27 Part B, 193-195.
- 727 [22] Körber, M., Gold, C., Lechner, D., Bengler, K., 2016. The influence of age on the take-over of
728 vehicle control in highly automated driving. *Transportation Research part F: Traffic*
729 *Psychology and Behaviour*. 39, 19–32.
- 730 [23] Gong J, Guo X, Pan L, Qi C, Wang Y., 2022. Impact of age on takeover behavior in automated
731 driving in complex traffic situations: a case study of Beijing, China. *Sustainability*. 14 (1), 483.
- 732 [24] Roche, F., Somieski, A., Brandenburg, S., 2019. Behavioral changes to repeated takeovers in
733 highly automated driving: effects of the takeover-request design and the nondriving-related
734 task modality. *Human Factors*. 61 (5), 839–849.
- 735 [25] Brandenburg, S., Roche, F., 2020. Behavioral changes to repeated takeovers in automated driving:
736 The drivers' ability to transfer knowledge and the effects of takeover request process.
737 *Transportation Research part F: traffic Psychology and Behaviour*. 73, 15-28
- 738 [26] Payre, W., Cestac, J., Delhomme, P., 2016. Fully automated driving: impact of trust and practice
739 on manual control recovery. *Human Factors*. 58 (2), 229–241.
- 740 [27] Neubauer, C., Matthews, G., Saxby, D., 2014. Fatigue in the automated vehicle: do games and
741 conversation distract or energize the driver? *Proceedings of the Human Factors and*
742 *Ergonomics Society Annual Meeting*. 58 (1), 2053–2057.
- 743 [28] Hongji Du, Xiaohua Zhao, Xingjian Zhang, Yunlong Zhang Jian Rong, 2015. Effects of fatigue
744 on driving performance under different roadway geometries: A simulator study, *Traffic Injury*
745 *Prevention*. 16 (5), 468-473.
- 746 [29] Li, S., Blythe, P., Guo, W. and Namdeo, A., 2018. Investigation of older driver's takeover
747 performance in highly automated vehicles in adverse weather conditions. *IET Intell. Transp.*
748 *Syst*. 12, 1157-1165.

- 749 [30] Gold, C., Körber, M., Lechner, D., Bengler, K., 2016. Taking over control from highly automated
750 vehicles in complex traffic situations: the role of traffic density. *Human Factors*. 58 (4), 642-
751 652.
- 752 [31] Zeeb, K., Buchner, A., Schrauf, M., 2015. What determines the take-over time? An integrated
753 model approach of driver take-over after automated driving. *Accident Analysis & Prevention*.
754 78, 212–221.
- 755 [32] Merat, N., Jamson, A. H., Lai, F. C. H., Carsten, O., 2012. Highly automated driving, secondary
756 task performance, and driver state. *Human Factors*. 54 (5), 762–771.
- 757 [33] Yoon, S. H., Ji, Y. G., 2019. Non-driving-related tasks, workload, and takeover performance in
758 highly automated driving contexts. *Transportation Research part F: Traffic Psychology and*
759 *Behaviour*. 60, 620–631.
- 760 [34] Naujoks, F., Purucker, C., Wiedemann, K., Marberger, C., 2019. Noncritical state transitions
761 during conditionally automated driving on German freeways: effects of non-driving related
762 tasks on takeover time and takeover quality. *Human Factors*. 61 (4), 596–613.
- 763 [35] Naujoks, F., Befelein, D., Wiedemann, K., Neukum, A., 2017. A review of non-driving-related
764 tasks used in studies on automated driving. *International Conference on Applied Human*
765 *Factors and Ergonomics*. pp. 525–537.
- 766 [36] Kim S, van Egmond R, Happee R., 2021. Effects of user interfaces on take-over performance: a
767 review of the empirical evidence. *Information*. 12 (4), 162.
- 768 [37] Kuehn, M., Vogelpohl, T., Vollrath, M., 2017. Takeover times in highly automated driving, (level
769 3). In *25th International technical conference on the enhanced safety of vehicles (ESV) national*
770 *highway traffic safety administration*, pp. 1-11.
- 771 [38] Yoon, S.H., Kim, Y.W., Ji, Y.G., 2019. The effects of takeover request modalities on highly
772 automated car control transitions. *Accident Analysis & Prevention*. 123, 150–158.

- 773 [39] Gold, C., Damböck, D., Lorenz, L., Bengler, K., 2013. “Take over!” How long does it take to get
774 the driver back into the loop? Proceedings of the Human Factors and Ergonomics Society
775 Annual Meeting. 57 (1), 1938–1942.
- 776 [40] Eriksson, A., Stanton, N. A., 2017. Takeover time in highly automated vehicles: noncritical
777 transitions to and from manual control. Human Factors. 59 (4), 689–705.
- 778 [41] Feldhütter, A., Gold, C., Schneider, S., Bengler, K., 2017. How the duration of automated driving
779 influences take-over performance and gaze behavior. In: Advances in ergonomic design of
780 systems, products and processes. Springer, Berlin, Heidelberg.
- 781 [42] Burrelly, A., Jacobé de Naurois, C., Zran, A., Rampillon, F., Vercher, J.-L., Bourdin, C., 2019.
782 Long automated driving phase affects take-over performance. IET Intell. Transp. Syst.. 13:
783 1249-1255.
- 784 [43] World Medical Association, 2013. World Medical Association Declaration of Helsinki: Ethical
785 principles for medical research involving human subjects. JAMA 310 (20), 2191–2194.
- 786 [44] Golding, J.F., 1998. Motion sickness susceptibility questionnaire revised and its relationship to
787 other forms of sickness. Brain Research Bulletin. 47 (5), 507-516.
- 788 [45] Johns. M.W., 1991., A New method for measuring daytime sleepiness: The Epworth Sleepiness
789 Scale. Sleep. 14 (6), 540–545.
- 790 [46] Smith, T.J., 2012. The attentional theory of cinematic continuity. Projections. 6 (1), 1-27.
791 Retrieved September 29th, 2022, from
792 <https://www.berghahnjournals.com/view/journals/projections/6/1/proj060102.xml>
- 793 [47] Dogan, E., Honnet, V., Masfrand, S., Guillaume, A., 2019. Effects of non-driving-related tasks
794 on takeover performance in different takeover situations in conditionally automated driving.
795 Transportation Research part F: Traffic Psychology and Behaviour. 62, 494-504
- 796 [48] Wierwille, W.W., Ellsworth, L.A., 1994. Evaluation of driver drowsiness by trained raters.
797 Accident Analysis & Prevention. 26 (5), 571-581.

- 798 [49] Belz, S. M., Robinson, G. S., Casali, J.G., 2001. An on-road investigation of commercial motor
799 vehicle operator self assessment of fatigue as an indicator of driver fatigue. Proceedings of the
800 Human Factors and Ergonomics Society Annual Meeting. 45 (23), 1576–1580.
- 801 [50] Yanko, M.R., Spalek, T.M., 2014. Driving with the wandering mind: the effect that mind-
802 wandering has on driving performance. *Human Factors*. 56 (2), 260–269.
- 803 [51] Jarosch, O., Gold, C., Naujoks, F., Wandtner, B., Marberger, C., Weidl, G., Schrauf, M., 2019.
804 The impact of non-driving related tasks on take-over performance in conditionally automated
805 driving – A review of the empirical evidence. 9. Tagung Automatisiertes Fahren
- 806 [52] Du, N., Zhou, F., Pulver, E., Tilbury, D.M., Robert, L.P., Pradhan, A.K., Yang, X.J., 2020.
807 Examining the effects of emotional valence and arousal on takeover performance in
808 conditionally automated driving. *Transportation Research part C: Emerging Technologies*.
809 112, 78-87.
- 810 [53] Russell, J.A., 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*.
811 39 (6), 1161–1178.
- 812 [54] Hergeth, S., Lorenz, L., Krems, J.F., 2017. Prior familiarization with takeover requests affects
813 drivers' takeover performance and automation trust. *Human Factors*. 59 (3), 457–470.
- 814 [55] Ebnali, M., Hulme, K., Ebnali-Heidari, A., Mazloumi, A., 2019. How does training effect users'
815 attitudes and skills needed for highly automated driving? *Transportation Research Part F:*
816 *Traffic Psychology and Behaviour*. 66, 184–195.
- 817 [56] Sibi, S., Balters, S., Fu, E., Strack, E.G., Steinert, M., Ju, W., 2020. Back to school: impact of
818 training on driver behavior and state in autonomous vehicles. In *2020 IEEE Intelligent Vehicles*
819 *Symposium (IV)*. IEEE Press, pp. 1189–1196.
- 820 [57] Yoon, S.E, Kim, Y.W. & Ji, Y.G. (2019) The effects of takeover request modalities on highly
821 automated car control transitions. *Accident Analysis & Prevention*, 123, 150-158)

- 822 [58] Tan, X. & Zhang, Y. (2022) The effects of takeover request lead time on drivers' situation
823 awareness for manually exiting from freeways: A web-based study on level 3 automated
824 vehicles, *Accident Analysis and prevention*, 168, 106593
- 825 [59] Davenne, D., Lericollais, R., Sagaspe, P., Taillard, J., Gauthier, A., Espié, S., Philip, P., 2012.
826 Reliability of simulator driving tool for evaluation of sleepiness, fatigue and driving
827 performance. *Accident, analysis & Prevention*. 45, 677–682.
- 828 [60] Engström, J., Johansson, E., & Östlund, J. (2005) Effects of visual and cognitive load in real and
829 simulated motorway driving. *Transportation research part F: traffic psychology and behaviour*,
830 8(2), 97-120
- 831 [61] Krause, M., Yilmaz, L., & Bengler, K. (2014). Comparison of real and simulated driving for a
832 static driving simulator. *Adv. Hum. Aspects Transp.: Part II*, 8, 29.
- 833 [62] Bougard, C., & Davenne, D. (2012). Effects of sleep deprivation and time-of-day on selected
834 physical abilities in off-road motorcycle riders. *European journal of applied physiology*,
835 112(1), 59-67.

Declaration of interests

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.

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.

Journal Pre-proof