

REVIEW

Open Access



Takeover performance evaluation using driving simulation: a systematic review and meta-analysis

Sónia Soares^{1†}, António Lobo^{1*†} , Sara Ferreira¹, Liliana Cunha² and António Couto¹

Abstract

Introduction: In a context of increasing automation of road transport, many researchers have been dedicated to analyse the risks and safety implications of resuming the manual control of a vehicle after a period of automated driving. This paper performs a systematic review about drivers' performance during takeover manoeuvres in driving simulator, a tool that is widely used in the evaluation of automated systems to reproduce risky situations that would not be possible to test in real roads.

Objectives: The main objectives are to provide a framework for the main strategies, experimental conditions and results obtained by takeover research using driving simulation, as well as to find whether different approaches may lead to different outcomes.

Methodology: First, a literature search following the PRISMA statement guidelines and checklist resulted in 36 relevant papers, which were described in detail according to the type of scenarios and takeover events, drivers' engagement in secondary tasks and the assessed takeover performance measures. Then, those papers were included in a meta-analysis combining PAM clustering and ANOVA techniques to find patterns among the experimental conditions and to determine if those patterns have influence on the observed takeover performance.

Conclusions: Less complex experiments without secondary task engagement and conducted in low-fidelity simulators are associated with lower takeover times and crash rates. The takeover time increases with the time budget of the first alert, which reduces the pressure for a driver's quick intervention.

Keywords: Automated vehicles, Takeover, Driving simulator, Systematic review, Meta-analysis

1 Introduction

Road crashes were responsible for more than 1.3 million deaths worldwide in 2016 [57]. In the European Union, road fatalities have been cut in half in around ten years, but still represented more than 27 000 lives lost in the same year [14]. Most crashes involving human costs are,

according to the European Commission (EC), directly or indirectly caused by human errors, such as distracted, fatigued, or drunk driving. To reduce the human role and eliminate this kind of crashes by 2050, the EC created the "Vision Zero" strategy [13], putting confidence in autonomous systems that should have safer driving capabilities than human driving. However, besides that a machine error is usually less accepted than a human error [34], even the best technologies may fail, especially when advances and adaptations occur every day during the ongoing transition to full automation. So far and for long years, automation will not completely replace human activity [25], imposing new coordination demands to the

*Correspondence: lobo@fe.up.pt

[†]Sónia Soares and António Lobo contributed equally to this work as first authors

¹ Research Centre for Territory, Transports and Environment, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal

Full list of author information is available at the end of the article

driver. Additionally, it is expected that automated vehicles will share the road environment with conventional, manually-driven vehicles. Therefore, beyond the technological progress, the rollout of automated vehicles on public roads requires intensive research on human factors. Moreover, it will be crucial to raise public awareness on the new risks introduced by the automated systems and to develop training actions about the appropriate use of automation, the prevention of risky behaviours, and the avoidance of misuse and disuse [42]. The research on automated driving has been exponentially growing in recent years to keep up with every step and ensure the safety of each new development.

1.1 Automation levels and the L2 versus L3 paradox

Automation levels are commonly represented by the SAE scale (L0-L5) [39]. The transition across different levels is an extensive and challenging process; notwithstanding, most of new cars already have some sort of driver assistance. L0 represents the absence of automated systems, being the driver completely and solely controlling the vehicle. L1 is very common and is characterized by having a single driver assistance system of either steering or acceleration/deceleration. The advanced driver assistance systems (ADAS), already present on L2, can be responsible for most of the dynamic driving task (DDT), allowing the driver to take the hands off the steering wheel and the feet off the pedals. However, the driver is still required to permanently monitor the road environment, being the only responsible for any failure that can occur. L3 automation not only performs all aspects of the DDT in some driving modes, but also is responsible for sending a takeover request (TOR) when the system faces a situation that it cannot handle. In such cases, the driver must be available to safely regain manual control, despite not being expected to constantly keep his/her eyes on the road while the system performs the DDT. This paradox has been hindering the homologation of L3 systems amid safety concerns, despite OEM's attempts to assure reasonable safety margins by capping L3 systems to specific traffic conditions at low speeds (e.g., traffic jam pilot). For this reason, the classification of an automated system as L2 or L3 can be merely a question of responsibility assumption instead of noticeable technological differences. From another perspective, L4 and L5 add fallback capabilities of the DDT, ensuring full automation for specific and for all driving environments, respectively. The predictable design of these two systems allows the driver to completely turn the mind off the DDT (e.g., by sleeping), or even the absence of a driver.

1.2 Automation failures and manual takeover

Nowadays, with only a few L2 vehicles on the market and the lack of legislation about L3 in most countries, a relatively long transition period is expected until L4 and L5 automated vehicles become common and affordable, which is expected not before 2040 [25]. Current automated systems are still far from being perfect and can fail for several reasons. Generally, the failures can be divided into two main groups: system limits and system malfunctions [7]. In the first case, the limitations are previously known and stated on the users' manual, meaning that they can be anticipated by the system itself, issuing a TOR, or by the driver that is observing the cues at the road environment. Examples of system limits are a traffic jam pilot that only works for specific ranges of speed and traffic volumes, a highway pilot that does not work in urban roads, or other limitations explicitly acknowledged by the manufacturer, such as difficulties in recognizing stationary vehicles or faded lane markings. System malfunctions raise higher concerns because the failure results from events unforeseen by the designers (e.g., algorithm errors or sensor breakdowns). This includes the inability to deal with certain situations (e.g., construction works and stationary objects) that were not tested or acknowledged as limitations [7]. As system malfunctions cannot be predicted by the driver, they may lead to imminent dangerous situations, although some systems may issue a real-time warning about a sensor failure or a sudden deactivation.

Apart from system failures, drivers' ability to react to unforeseen situations cannot be ignored, especially when referring to L3 and lower automation levels. As the automation progresses, the role of human drivers becomes less and less active, but even at lower levels such as L1 and L2, drivers' situation awareness tends to decrease. This can happen for the simple fact that the physical disengagement of the DDT causes boredom or fatigue and because the facilitated driving better allows the engagement in non-driving-related tasks (NDRTs) [11, 42]. With or without a TOR, drivers must be prepared to act when necessary. The alertness and promptness to resume the manual control are important factors affecting the effectiveness and safety of the intervention. Drivers' abilities rely on many factors, such as age, gender, driving experience, or drowsiness state [43, 44], but driving conditions, such as the environment complexity, traffic density, visibility, or time available to safely react (time budget), have a fundamental role in the takeover performance [19, 40].

1.3 Research motivation and objectives

As fully autonomous vehicles will not be available overnight, the shared control of dynamic driving between human and machine is the main challenge for the

advancement of vehicle technology in a context of increasing automation [11, 23]. In particular, takeover performance is the main safety concern related with partial (L2) and conditional (L3) automation. In recent years, numerous studies evaluated takeover performance through time and quality indicators that characterize drivers' reactions [1, 63]. Most of this research has been conducted in virtual environment for ethical and practical reasons related with the safety of participants and the lack of widespread technologies and infrastructures for real-world testing of automated vehicles.

The large body of literature on takeover during automated driving has sparked motivation for review studies that aggregate and summarize the knowledge obtained through a large spectrum of experimental conditions reflected in multiple takeover scenarios and diverse apparatus and participants' characteristics. Radlmayr and Bengler [36], Vogelpohl et al. [51], and Walch et al. [52] are examples of earlier reviews of takeover time and/or quality studies, developing a narrative analysis of studies published before 2016. Eriksson and Stanton [12] conducted a quantitative review of takeover time, showing that this variable is positively correlated with the time budget. However, the authors did not extract from the literature any other effects that may affect the takeover time. More recently, Zhang et al. [65] acknowledged the lack of quantitative studies and presented one of the most comprehensive reviews of takeover times to date. The authors used a within-study analysis, a between-study analysis, and a linear mixed-effects model. The results showed that shorter takeover times are associated with urgent takeover events, not using handheld devices, not engaging in visual NDRTs, having experienced a previous takeover event during the experiment, and receiving an auditory or vibrotactile TOR. However, Zhang et al. [65] did not address takeover quality in their review.

Similarly to Zhang et al. [65], the present review is also motivated by the few existing syntheses of takeover time and quality that follow a quantitative approach, as well as by the need to consider recent studies in a fast-moving research field. However, rather than investigating the individual effects of different variables on takeover performance measures, the approach followed in this study is unique in the sense that: (i) the meta-analysis is focused on mining patterns among diverse experimental conditions used in takeover research to understand whether those patterns may be associated with different outcomes, and (ii) a comprehensive narrative review contextualizes the most relevant experimental conditions, supports the interpretation of the results from the quantitative analysis, and provides guidance for future research automated driving research. For the sake of consistency, this review is limited to driving simulator

studies. In this context, this study aims to address the following questions:

- Which experimental conditions and simulated scenarios have been used to study takeover?
- Do the different experimental conditions play an important role on the outcomes of takeover performance evaluation?

To accomplish these objectives, the methodology adopted in this review included three main steps. First, a systematic review of existing literature was conducted following the PRISMA statement guidelines and checklist [30]. Second, a descriptive analysis focused on the selected papers allowed to synthesize the experimental conditions, including the simulation scenarios, NDRT engagement, presence of TOR, and takeover performance measures. Finally, the quantitative analysis used PAM clustering and ANOVA techniques to explore different patterns characterized by takeover performance measures, simulation conditions, driver characteristics, and publication rankings.

2 Search methods

The review started with a deep search in target databases following the PRISMA statement guidelines and checklist [30]. The databases selected for this search were the Web of Science (WoS), Scopus, and Transportation Research International Documentation (TRID), considering that these databases include a wide range of relevant papers in the transportation field. The search was performed in the title, abstract, and keywords of the papers indexed to the databases, using the following combination of terms: (autonomous OR automated OR self-driving) AND (driving OR car OR vehicle) AND (driver simulator OR driving simulator). Two filters were applied to limit the search to documents written in English and published between January 2015 and May 2020. Studies published before 2015 were excluded to ensure that the selected studies are up-to-date with the fast development of automated driving paradigms and technology observed in the last few years, without detriment to the consideration of the large majority of the existing studies in the field. Through this selection procedure, 249 papers were obtained from WoS, 173 from Scopus, and 172 from TRID, resulting in 594 documents to scan. Additional research identified though other sources resulted in one complementary relevant paper. After removing the duplicates, this number was reduced to 370 papers. Given the large number of documents, a filtering procedure was fundamental to be more insightful on the quality of the selected papers. Only the papers published in journals indexed to WoS,

i.e., with Journal Impact Factor, were retained, which reduced the sample to 255 documents. Then, the remaining documents were screened by analysing the titles and abstracts, which led to the exclusion of 210 papers according to the following criteria:

- Research unrelated to automated driving (e.g., medicine, robotics, augmented reality, traffic optimization, railways);
- Focused on bicyclists, motorcyclists, and truck platooning;
- Focused on sustainability and eco-driving;
- Focused on trust, comfort, architecture, and design;
- Automation level too low or inexistent (e.g., a system only working with advanced cruise control);
- Not essentially sustained on driving simulator experiments;
- Studies going deep into systems components, computer science and complex algorithms.

The 45 papers that passed through the screening process were read and examined in depth to assess if there were any reasons to exclude more studies from the analysis. Considering the previous research questions, it was decided to exclude nine studies that did not report takeover time or quality measures. All the remaining 36 studies were included in the analysis and report takeover times, considered as the elapsed time between a TOR issued by the vehicle and the first manual input in the steering wheel or pedals, or, in the absence of a TOR, to the elapsed time between the moment when a potential danger becomes visually apparent in the simulation scenario and the first manual input. The time-to-collision (TTC) and the crash rate are the other relevant takeover performance measures reported in the 36 selected papers, albeit with a much smaller incidence: the TTC is reported in nine studies and the crash rate in just eight. The literature search and selection procedure is summarized by the flow diagram in Fig. 1.

The selected papers are the following: Alrefaie et al. [1], Blommer et al. [2], Bourrelly et al. [3], Choi et al. [4], De Winter et al. [6], DeGuzman et al. [7], DinparastDjadid et al. [8], Dogan et al. [9, 10], Feldhütter et al. [15], Gold et al. [17], Happee et al. [18], Lin et al. [24], Lodinger and DeLucia [26], Louw et al. [27, 28], Lu et al. [29], Naujoks et al. [31], Payre et al. [32, 33], Rauffet et al. [37], Shen and Neyens [41], Vogelpohl et al. [48–50], Wan and Wu [53], Wandtner et al. [54, 55], Wu et al. [58–60], Yoon and Ji [61], Zeeb et al. [62–64], Zhou et al. [66]. These papers were published in the following journals: *Accident Analysis & Prevention*, *Applied Ergonomics*, *Cognition*, *Technology & Work*, *Human Factors*, *IEEE Access*, *IEEE Transactions on Human–Machine Systems*, *IET*

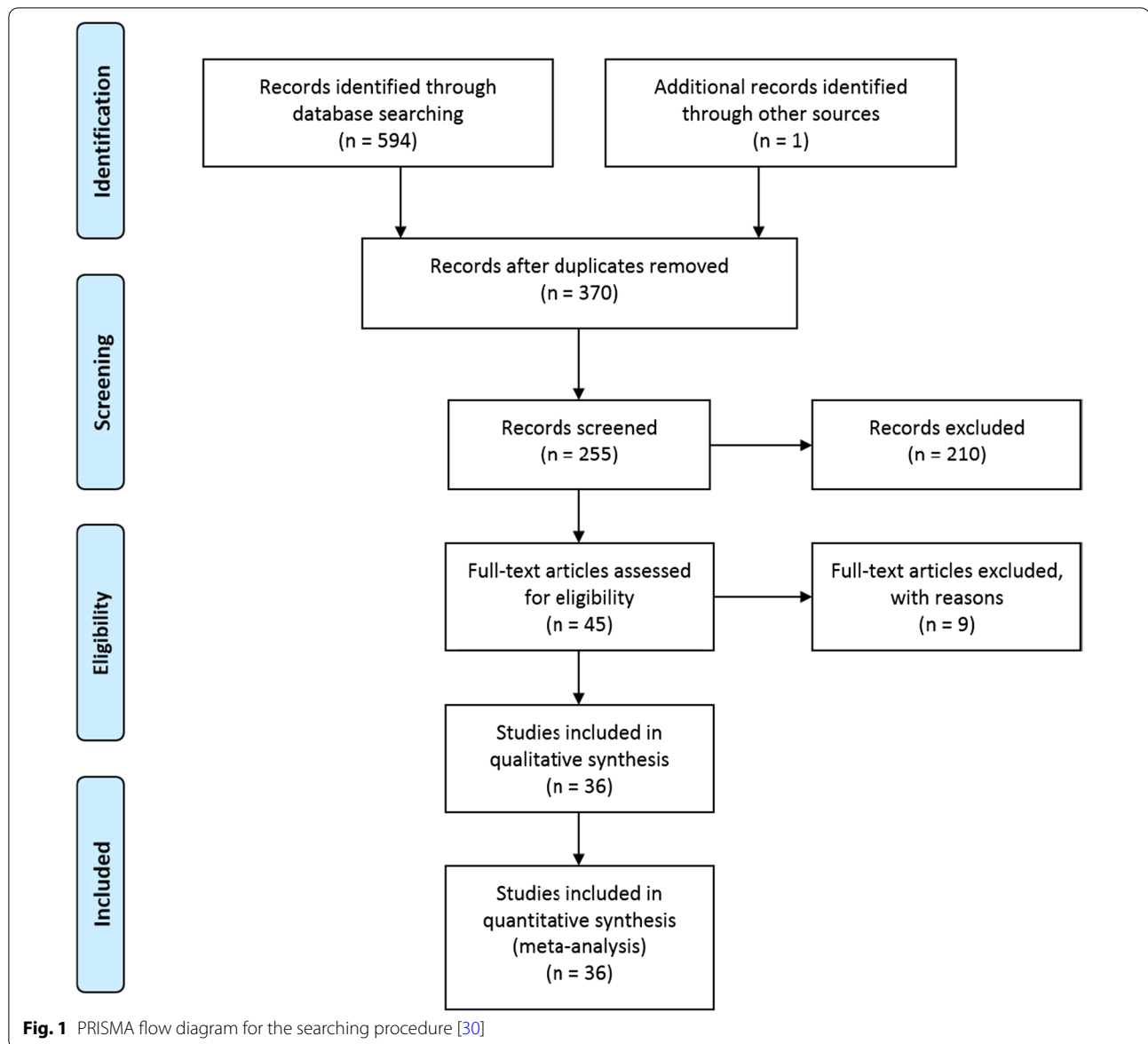
Intelligent Transport Systems, *International Journal of Vehicle Design*, *Journal of Safety Research*, *Transportation Research Part F: Traffic Psychology and Behaviour*, and *Transportation Research Record*. The studies by Gold et al. [17] and Happee et al. [18] use data from previous works performed by the authors.

3 Descriptive analysis

The experimental settings related to the evaluation of takeover performance in driving simulator do not follow a single protocol, resulting in very distinct procedures among experiments. However, some common points can be found. First, a pre-experiment driving task is described in almost all studies. This training phase is fundamental in every driving simulation experiment to produce reliable results, by alleviating the learning effects. For automated driving simulations, this training was even more important since most people have never experienced higher levels of automation. In this sense, most studies provided a few minutes for drivers' familiarization, usually including transitions from manual to automated driving and vice-versa.

Concerning automation levels, the study of takeover safety is always focused on L2 (“hands-off”/partial automation) or L3 (“eyes-off”/conditional automation), featuring simultaneous speed and steering assistance, or traffic jam assistance. When the automation level is not explicitly mentioned, L2 and L3 can be distinguished by the need for a constant monitoring of the road, which is mandatory in L2 [39]. L4 automation (“mind off”) is capable of bringing the vehicle to a stop at a safe location if the driver does not assume the manual control when the road environment is out of the limits of technology, therefore takeover safety is not a problem. In L5, only voluntary takeovers may occur.

The following sections describe the experimental conditions that vary significantly across different studies. These conditions are related with the engagement in NDRTs, the simulated takeover events, the presence of a TOR, and the analysed takeover performance measures. The characteristics of the participants in the experiments are not under the scope of this analysis. In fact, it was detected that many driving simulation studies on takeover use convenience samples, being focused on the effects of different driving scenarios rather than on the participants' features. Nevertheless, generic characteristics of the participants' sample considered in each paper are summarized in Additional file 1: Appendix, together with the type of experimental design, the most relevant takeover performance measures and the main findings.



3.1 NDRT engagement

When conducting a literature review about automated vehicles, approaching the issues related with NDRTs is unavoidable. Besides the safety reasons associated with the EC's "Vision Zero" strategy [13], the evolution of automated vehicles brings practical benefits for the users that become allowed to engage in other tasks besides de DDT. Driving is time-consuming and can be boring and stressful [5], thus automation will allow drivers to combine their trips with work, rest, or leisure. The 36 reviewed studies do not neglect this fact. From those, 15 had the presence of a distractive task, five were performed without it, 17 had a combination of both situations, and two did not provide such information. In most

of the 29 studies with NDRTs, the engagement on a specific activity was mandatory. This guaranteed that drivers were all at the same conditions during the takeover and allowed to analyse the effects of distinct tasks and to compare them with a baseline scenario without distraction. To ensure drivers' engagement on the NDRT, some studies incentivized the participants to dedicate themselves to the activity, assuming it as primordial (e.g., exchange as many e-mails as possible, respond to questions, or get points in a game). However, five studies allowed for an optional task engagement at least in some experiments [5, 10, 15, 29, 31], with the driver being allowed to interrupt the task or to perform it only when he/she felt safe.

NDRTs can be classified into 4 main categories: visual, auditory, motor, and cognitive. However, these stimuli are not usually present in an isolated form. For instance, playing a game can combine the four types: (i) looking at the playing device involves visual distraction, (ii) every game involves at least a selection of an answer or touching a virtual object, i.e., motor distraction, (iii) if the game is a trivia or needs concentration, a cognitive distraction is involved, and (iv) if the game is played with sound, an auditory distraction is present too. To illustrate the distractive domain under investigation, Table 1 summarises the NDRTs identified in the selected studies.

The visual distraction was the most present among the experiments, and playing a game was the most reproduced task, being present on 16 studies. In fact, playing a game is a very complex activity that can involve all distraction types and requires drivers' full engagement. Examples of used games are *Tetris*, *Sonic Dash*, *Angry Birds*, *Candy Crush*, trivia, labyrinths, anagrams, and specific tasks created by the researchers, such as connecting dots or identifying objects. Some studies followed the methods to assess driver demand due to the use of in-vehicle systems defined in ISO/TS 14198, standard from 2012 revised in 2019 [20, 21], namely by introducing a surrogate reference task (SuRT) or an n-back task. SuRT requires locating target stimuli on a screen (visual and motor distraction), and was introduced in four studies

[2, 4, 17, 18]. Because of the similarities of induced stimuli, SuRT was included in the "play a game" category. N-back is an auditory-vocal task that requires memorizing sequences of numbers (cognitive distraction), being used in five studies (Table 2). After gaming, watching a video and reading were the most common activities used to distract drivers in driving simulation studies.

Secondary task engagement is normally seen as a problem that may delay the manual control recovery in a critical situation, with many studies confirming such concerns. For instance, Wandtner et al. [54] introduced an auditory-vocal, a visual-vocal, and two visual-manual NDRTs to demonstrate that drivers' responses and perceived safety were strongly affected by NDRTs. Overall, perceived safety (based on a self-assessment scale) decreased while braking response time increased in the following order: no task, auditory-vocal, visual-vocal, and visual-motor. Differences were observed even between the two visual-motor tasks: performing a task with a handheld device had stronger effects than performing it with the device mounted on the dashboard. In the same line, Zeeb et al. [64] observed larger reaction times for drivers interacting with a handheld tablet than with a tablet fixed on the dashboard. Dogan et al. [9] and Wu et al. [58] found that the minimum TTC tends to decrease when drivers engage in NDRTs under automated driving in relation to manual driving, which poses an increased

Table 1 NDRTs identified in the selected studies

NDRT	Involved stimuli				References
	Visual	Auditory	Motor	Cognitive	
E-mail	●		●	(●)	Alrefaie et al. [1], Dogan et al. [9], Zeeb et al. [63]
20 questions task		●		●	Alrefaie et al. [1], Lodinger and de Lucia [26]
Adjust cabin temperature	●		●		Blommer et al. [2]
Search for a song, radio station or internet search	●		●		DeGuzman et al. [7], Yoon and Ji [61], Zeeb et al. [62]
Watch a video	●	●			Bourrelly et al. [3], Dogan et al. [9], Feldhütter et al. [15], Lin et al. [24], Shen and Neyens [41], Wan and Wu [53], Wu et al. [59], Yoon and Ji [61], Zeeb et al. [63]
Use mobile phone	●		●		Dogan et al. [10], Yoon and Ji [62]
Read	●		(●)		Dogan et al. [7], Feldhütter et al. [15], Lin et al. [24], Vogelpohl et al. [50], Wan and Wu [53], Wandtner et al. [54], Zeeb et al. [63, 64]
Play a game	●	(●)	●	(●)	Blommer et al. [2], Choi et al. [4], De Winter et al. [6], DinparastDjadid et al. [8], Feldhütter et al. [15], Gold et al. [17], Happee et al. [18], Louw et al. [27, 28], Lu et al. [29], Payre et al. [32], Rauffet et al. [37], Vogelpohl et al. [50], Wan and Wu [53], Wu et al. [58, 59]
N-back task		●		●	Choi et al. [4], Gold et al. [17], Happee et al. [18], Louw et al. [28], Wu et al. [58]
Office work	●		●	(●)	Naujoks et al. [31], Vogelpohl et al. [48]
Typing or texting	●		●		Gold et al. [17], Wan and Wu [53], Zeeb et al. [62]
Take a nap	●	●	●	●	Wan and Wu [53]
Repeat sentences		●		●	Wandtner et al. [54]
Transcribe sentences	●		●		Wandtner et al. [54, 55], Zeeb et al. [62, 64]

Legend: ● the NDRT involves this type of distraction; (●) in some experiments, the NDRT may involve this type of distraction

Table 2 Takeover scenarios identified in the selected studies

Takeover event	References
Obstruction of the ego-vehicle's lane, roadwork zone or lane reduction	Blommer et al. [2], Choi et al. [4], DinparastDjadid et al. [8], Dogan et al. [9], Feldhütter et al. [15], Gold et al. [17], Happee et al. [18], Lin et al. [24], Rauffet et al. [37], Vogelpohl et al. [48, 50], Wan and Wu [53], Wandtner et al. [54], Wu et al. [58–60], Yoon and Ji [61], Zeeb et al. [62, 63], Zhou et al. [66]
Lead-vehicle deceleration/braking or slow lead-vehicle	Choi et al. [4], DinparastDjadid et al. [8], Lodinger and DeLucia [26], Louw et al. [28], Vogelpohl, et al. [48–50]
Vehicle cuts into the ego-vehicle's lane	Lodinger and DeLucia [26], Naujoks et al. [31], Payre et al. [33], Zeeb et al. [64]
Missing/faded lane markings	Dogan et al. [9], Naujoks et al. [31], Payre et al. [32], Vogelpohl et al. [48–50], Zeeb et al. [63, 64], Zhou et al. [66]
Lane opening or diverging lanes	DeGuzman et al. [7], Yoon and Ji [61], Zeeb et al. [63]
Pedestrian/dog/cyclist crossing	De Winter et al. [6], DinparastDjadid et al. [8], Lu et al. [29]
Wind gust blowing	Shen and Neyens [41], Zeeb et al. [63, 64]
Limited visibility due to adverse weather	Lodinger and DeLucia [26], Louw, et al. [28], Vogelpohl, et al. [49, 50], Zhou et al. [66]
Change of road environment or traffic conditions	Dogan et al. [10], Wandtner et al. [55]
Sensor failure or system's shut-down	DeGuzman et al. [7], DinparastDjadid et al. [8], Naujoks et al. [31], Louw et al. [29], Payre et al. [32, 33]

safety risk. Wan and Wu [53] was the only study analysing the possibility of a driver being asleep at the time of an urgent TOR, simulating a situation that would configure a misuse of L2 and L3 automated systems.

Nevertheless, some authors recognize that a secondary task may be important to avoid drivers' fatigue and boredom. For instance, Wu et al. [59] investigated the effects of NDRTs on drowsiness, focusing on finding differences between distinct age groups. The results showed that for younger drivers, engaging on NDRTs was beneficial in combating drowsiness and did not disturb takeover performance. On the other hand, task engagement in older drivers did not affect drowsiness development but degraded takeover performance, especially under more complex tasks. Middle-aged drivers fell at an intermediate level between the other two groups. Gold et al. [17] also confirmed the benefits of at least certain tasks to improve takeover performance by increasing the minimum TTC, but Feldhütter et al. [15] did not find relevant differences in this variable. In a different perspective, Blommer et al. [2] used mandatory secondary tasks to avoid drivers' fatigue and boredom, but ensured that such tasks were performed only during automated driving and ceased prior to the critical events.

3.2 Takeover events

This section focuses on describing the takeover events created for the experimental studies. As mentioned above, automation failures that lead to manual control recovery can occur due to known and predictable system boundaries (system-limit failure), or malfunctions unforeseen by the system's designers (system-malfunction failure). A recent study conducted by DeGuzman

et al. [7] analysed the takeover performance of two groups of participants, one experiencing a system-limit failure, and the other a system-malfunction failure. The differences found in takeover behaviour between the two groups were clear. The system-limit failure was associated with drivers more prepared to act, revealing a higher percentage of time looking at the roadway and in-vehicle cues before the failure occurs, as well as smaller takeover times.

In addition to the type of failure, the event urgency also represents a factor that may affect takeover performance. When the driver needs to recover the manual control of the vehicle, there is a situation that automation cannot handle, regardless if it results from a system's limitation or malfunction. These situations can assume many forms, including the need to perform a specific manoeuvre, the need to correct an action (or inaction) of the system, or the variation of road and traffic conditions, and may or not present a risky situation. To illustrate this, Vogelpohl et al. [50] tested two takeover scenarios: missing lane markings and roadworks. Both scenarios were designed with low and high urgency, respectively, resulting in distinct visual patterns. The low-urgency situation was associated with shorter times between the TOR and the first gaze at the speedometer, while the high-urgency event resulted in shorter times for the first gaze at the road centre, hands on the steering wheel, and the first gaze at the side mirror.

Regardless the urgency of the situation recreated in simulated environment, most of the analysed takeover events had an associated collision risk if no human

intervention occurred. In this context, the simulated hazards are listed in Table 2.

The obstruction of the ego-vehicle's lane was the most replicated takeover event, being identified in 20 studies. Obstacle avoidance is a very common manoeuvre that can occur in real driving for several reasons, e.g., a broken vehicle [62], construction works [8], or a simple lane reduction [66]. The virtual objects blocking the lane represented on the experiments included stationary vehicles (with or without flashing warning lights), crashing scenes, police vehicles, falling objects, and traffic signs/cones. The presence of a slow lead vehicle or the sudden breaking of this vehicle are situations that share some similarities with lane obstruction, obliging the ego vehicle to break or to perform an avoidance manoeuvre. These situations were simulated in seven studies.

Following obstacle avoidance, the failure to detect the lane markings was the most reproduced event. An adequate performance of an automated vehicle depends on a full understanding of the road layout, which is mostly provided by lane markings [45]. If the vehicle is not able to detect lines or arrows, driving safety can be compromised, and a TOR is the most reliable option to deal with such situation. These events are simulated by faded lines, poor visibility, or leaves covering the lines. Adverse weather, such as rain, snow or fog, is used to simulate the limitations of an automated system under low visibility, but its effects are mostly combined with other occurrences, such as lane obstruction, which does not allow to generalize conclusions. Nevertheless, two studies have assessed weather effects in a more comprehensive way, albeit obtaining contradictory results. Louw et al. [28] found a degraded takeover performance in scenarios with fog, attributing this fact to the low visibility, but Vogel-pohl et al. [50] showed quicker reactions under rainfall, which may be associated with drivers' increased efforts to stay alert and monitor the road environment.

Other important category of simulated events is the shut-down of the automated system, observed in 6 studies. In some cases, the failure may be accompanied by a TOR that warned for the deactivation of the system [8, 31–33], or be a “silent” failure leading to subtle [27] or sudden [7] trajectory deviations.

Other less represented situations included other vehicles cutting into the ego vehicle's lane, the opening of new lanes, the crossing of vulnerable road users and the change of road/traffic conditions. Table 1 also shows that several studies analysed multiple events, but Alrefaie et al. [1] and Bourrelly et al. [3] are not included. Alrefaie et al. [1] simulates critical events but do not provide details that allow to classify them. Bourrelly et al. [3] focused exclusively on the effects on driving performance after a long period of automation, issuing a TOR without a specific cause or visual cue. The authors concluded that the longer the driving period, the longer the reaction times and the sharper the avoidance manoeuvres. They also suggested that frequent TORs could improve takeover performance, since taking over control seemed to eliminate the impact of the accumulated passive fatigue. In the same vein, Naujoks et al. [31] concluded that drowsiness affects drivers' performance in partially automated driving, while low to moderate levels of visual and mental workload improved the performance in the riskiest driving scenarios. Feldhütter et al. [15] showed another perspective and did not find significant differences in the takeover performance between two conditions, in which one should be associated with higher fatigue levels than the other. However, the authors justified that short relaxing moments, or even sleeping periods, had a certain reviving effect, so that the drivers could be no longer fatigued during the takeover event.

From the analysed studies, the great majority were conducted in medium-(11 studies) or high-fidelity simulators (17 studies). Both types feature a real or mock-up car and immersive video projection, with the latter adding dynamic feedback capabilities. Five studies were conducted in low-fidelity simulators, consisting of a gaming steering wheel and pedals, regular monitors and, sometimes, a car seat [24, 26, 53, 61, 66]. One study combined experiments in low- and medium-fidelity simulators [6]. Gold et al. [17] and Shen and Neyens [41] did not provide information about the characteristics of the simulators.

Table 3 Presence of TOR in the selected studies

TOR	References
Yes	Alrefaie et al. [1], Blommer et al. [2], Bourrelly et al. [3], Choi et al. [4], DinparastDjadid et al. [8], Dogan et al. [9, 10], Feldhütter et al. [15], Gold et al. [17], Happee et al. [18], Lin et al. [24], Lu et al. [29], Naujoks et al. [31], Payre et al. [32, 33], Rauffet et al. [37], Vogel-pohl et al. [48–50], Wan and Wu [53], Wandtner et al. [54, 55], Wu et al. [58–60], Yoon and Ji [61], Zeeb et al. [62–64]
No	De Winter et al. [6], DeGuzman et al. [7], Lodinger and DeLucia [26], Louw et al. [27, 28], Shen and Neyens [41]
Some experiments	Zhou et al. [66]

3.3 Presence of TOR

The need for resuming the manual control of the vehicle after a period of automated driving was present in every selected study. However, a driver's consciousness about this need may be triggered by a TOR issued by an in-vehicle warning system or by the visual perception of potential hazards or critical events. The majority of the studies (29) included in the review provided a request before the system limits were reached, while six studies did not provide such alert, considering it as a "silent" failure [27]. Only one study considered both situations. Table 3 presents the studies that implemented or not a TOR in the driving simulations.

Moreover, human factors affecting the interaction between drivers and automated systems, such as the level of acceptance, trust and reliance on technology, together with drivers' knowledge about the system limits, may have strong impacts on the decision to take over control. These factors not only impact the promptness of drivers' reaction to a TOR, consequently affecting takeover performance, but also impact the earlier decision to undertake an NDRT. In some studies, the authors opted to provide all the information about how the system works and its limitations and boundaries [6, 7, 26]. From another perspective, some studies aimed to analyse a more critical scenario and do not inform participants that automation would fail [27, 41]. This background on the system's functionalities is essential to define a driver's level of alertness and interaction with environmental cues. It is logical to assume that, if a driver trusts on automated technology and/or is told that the system will emit an anticipated alert when human intervention is needed, he/she will be more relaxed and less engaged in the monitoring task. However, as the risk perception in simulated environment is much lower than in real-world driving, many of the reviewed studies do not approach these issues, lacking information about the participants' trust and previous knowledge on the performance of automated systems.

3.4 Takeover performance measures

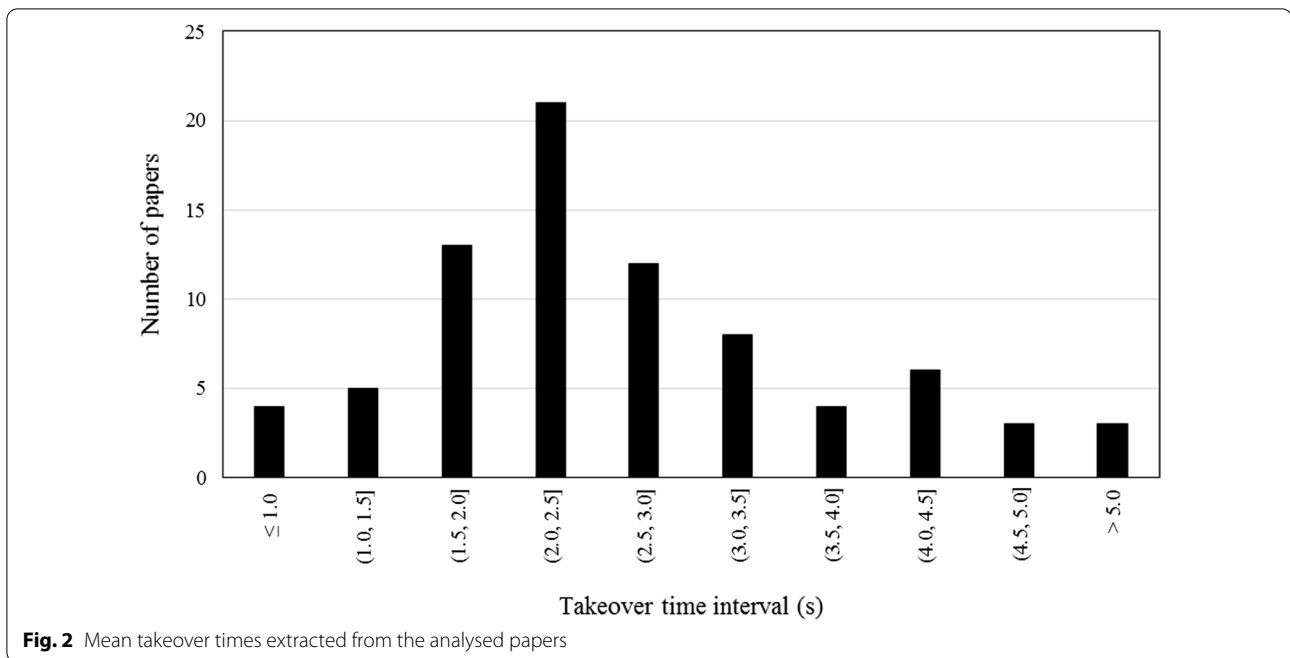
To evaluate the efficiency of drivers' manual inputs, studies observed indicators of takeover quality. The most common variables are the takeover time (also known as response or reaction time), TTC, crash rate, and variations in driving dynamic parameters (see the Additional file 1: Appendix).

The definition of reaction time is not a global and well-defined concept, and there are a variety of measures, depending on the event that generates the response [16]. In autonomous driving, a commonly adopted definition for takeover time is the time between the takeover stimulus and the moment of driver intervention [65]. However,

this definition is too generic and questions are raised about what is considered as the first stimulus and how a driver's intervention can be verified. Regarding the first stimulus, commonly the takeover time is measured as the elapsed time since the first auditory, visual, and/or haptic TOR [4, 48, 64]. Nonetheless, some studies calculate the takeover time without a TOR, for instance considering the first visual cue of a hazard or the start of an untypical behaviour from the automated system (e.g., due to a malfunction) [7, 27]. Concerning driver intervention, studies generally assume the intervention as the system deactivation, which, depending on the system, can occur due to the first pedal or steering wheel input or by pressing a button. Nonetheless, some studies specify the boundaries of assuming a steering or braking reaction. For instance, the steering reaction was considered as a minimum turning of 2 degrees by Louw et al. [28] and Wan and Wu [53]. Lodinger and de Lucia [26] considered a brake response as 2% of the maximum possible brake pressure, and Wan and Wu [53] and Zeeb et al. [62] increased that limit to 10%. In the analysed studies, the mean takeover time was extracted as the smallest value between braking, steering, or system deactivation, depending on the available information. The time until the first gaze on the hazard and the time until putting the hands on the wheel was also calculated in some studies, however, to have homogeneous data, those times were not considered in this review. Figure 2 shows the number of papers presenting mean takeover times within intervals of 0.5 s.

All the 36 selected papers reported takeover times, and each paper may report more than one mean takeover time value to represent different experimental conditions and/or participants characteristics. In total, 150 mean takeover time values were extracted from the papers, 75% of which fall between 1.5 and 3.5 s.

In relation to the TTC, Feldhütter et al. [15] defined it as "the theoretically remaining time until a potential collision with an obstacle assuming a constant speed of the ego-vehicle", which means that higher TTCs should represent safer behaviours [3]. However, as previously mentioned, some authors analysed NDRT engagement in relation to the minimum TTC with either positive or negative effects [9, 15, 17, 58]. The TTC was also used for comparison between visual and auditory tasks [54], partial and highly automated driving [31], and short and long automated driving periods [3, 60]. Happee et al. [18] focused on the analysis of different TTC measures. In total, only 9 studies measured the TTC. In turn, 19 studies refer to the time budget, defined as the time for a system's deactivation or as the TTC at the moment a TOR is issued due to a critical event, corresponding in this case approximately to the sum of the takeover time with the minimum TTC. Because of both definitions, the time



budget varies over a wide range (1.7–60.0 s) in the sample studies, being less relevant for the assessment of takeover performance than the takeover time or the TTC. For instance, with the reduction of the takeover urgency, Wan and Wu [53] found that the probability of having a crash stabilizes at near zero for time budgets above 10 s.

The number of collisions, or crash rate, is also an output related with driver behaviour, as it is common sense that poorer driving performance increases crash probability. For instance, Williamson et al. [56] conducted experiments to study the effects of drowsiness on manual driving performance and results showed that the majority of participants reporting increasing sleepiness levels had a higher likelihood of crashing. In the case of imminent risks, which are often the cause of TORs under automated driving, a timely reaction should be accompanied by the capability of avoiding that risk. For this reason, the crash rate is an important measure of takeover performance in pair with time measures. However, takeover quality has been much less analysed than takeover time. Taking a look at the 36 selected papers, only eight of them reported crash figures in the conducted experiments. From those, Choi et al. [4], De Winter et al. [6], Lodinger and DeLucia [26], and Naujoks et al. [31] reported zero crashes. The remaining studies obtained collision rates varying from 0 to 60%, depending on the time budget, the NDRT engagement and the type of NDRT, and gaze behaviour. Zeeb et al. [62] categorized their sample into “high”, “medium” and “low-risk” drivers, according to the percent time looking at the road and gaze behaviour, and

found that high-risk drivers collided significantly more often (45.0%) than low-risk drivers (15.2%). Lin et al. [24], Wan and Wu [53], and Wandtner et al. [54] found that drivers distracted with NDRTs with increasing difficulty collided more often and simultaneously presented higher takeover times. Therefore, while secondary tasks may help to reduce drowsiness [59], they also seem to pose an additional crash risk under critical takeover situations.

Besides crash figures, takeover quality can be reflected in dynamic driving variables. The most common indicators found in this review were changes in the acceleration and lateral position. Louw et al. [28] stated that “steering collision avoidance remains a feasible option for a longer time than braking avoidance, during the run-up to a potential collision”, with 55% of their sample mainly steered in response to a lead vehicle. Wandtner et al. [55] used standard deviation of lateral position (SDLP) and percentage of lane excursions to assess lateral control quality when recovering manual control. Both SDLP and lane excursions revealed a performance decrement when drivers were engaged in secondary tasks. The authors also analysed vehicles’ lateral and longitudinal accelerations, and results showed that during the NDRT, drivers generated significantly higher maximum accelerations. Alrefaie et al. [1] considered the mean percentage change of vehicle’s speed and heading angle for a period before takeover, associating higher values with poorer takeover performance. Lane excursions were analysed by Louw et al. [27], and the effects of NDRTs were significant (30% of lane excursions occurred in normal conditions,

while 70% occurred for drivers distracted with non-driving activities). Moreover, lane excursions occurred more often on curved road sections than on straight road sections.

4 Quantitative analysis

A meta-analysis was performed to find common patterns among the reviewed studies and to investigate how the experimental conditions in those studies may have influenced the measured outcomes. Generically, all the 36 studies had the objective of analysing takeover performance, thus the most frequent time and quality measures, i.e., takeover reaction time and crash rate, were chosen as target variables.

4.1 Data description

The data used for this analysis was extracted directly from the selected papers. As previously mentioned, most authors performed different experiments to simulate different takeover conditions. In this sense, the authors presented the results of takeover performance by averaging takeover time and quality indicators according to different aggregation criteria, such as the type of takeover event, the type of NDRT, or some participants' characteristics. In total, 150 values of mean takeover time (TOT) were extracted from the studies, of which just 22 were associated with the corresponding mean crash rate (CR).

These 150 observations were completed with information about the following variables:

- Simulator fidelity (FID): low, medium, or high;
- NDRT engagement: no, optional, or mandatory;
- Presence of TOR: no or yes;
- Mean time budget (TB);
- Participants' sample size (SAMPLE);
- Participants' mean age (AGE);
- Percentage of male participants (MALE);
- Journal's highest quartile in WoS's 2020 Journal Citation Reports (QUART): Q1, Q2, and Q3/Q4.

All the variables were selected by seeking homogenous data across the different sources. Nevertheless, in some cases the information is not available. The inclusion of sample characteristics aims to enrich the quantitative analysis, despite being out of the scope of the descriptive analysis, focused on the simulation characteristics. The journal ranking represents a standardised measure of its quality. The 3rd and 4th quartiles were grouped under the same category for practical reasons, because there was only one paper on a Q3 journal and four papers on Q4 journals. The data description is presented in Table 4.

Table 4 Data description for the meta-analysis

Variable	Mean	Standard deviation	Minimum	Maximum	Relative frequency (%)
TOT (s)	2.6	1.0	0.4	5.2	–
CR (%)	7.6	11.9	0.0	45.0	–
FID=low	–	–	–	–	13.3
FID=medium	–	–	–	–	28.0
FID=high	–	–	–	–	52.7
FID=n/a	–	–	–	–	6.0
NDRT=no	–	–	–	–	25.3
NDRT=optional	–	–	–	–	6.7
NDRT=mandatory	–	–	–	–	48.0
NDRT=n/a	–	–	–	–	20.0
TOR=no	–	–	–	–	16.0
TOR=yes	–	–	–	–	84.0
TB (s)	6.7	2.6	1.7	10.0	–
Sample	31.3	16.3	6.0	84.0	–
Age (years)	36.8	9.3	19.8	64.3	–
Male (%)	54.4	12.4	17.0	100.0	–
QUART=Q1	–	–	–	–	38.7
QUART=Q2	–	–	–	–	50.7
QUART=Q3/Q4	–	–	–	–	10.7

4.2 Methods

To find patterns associating the experimental conditions (FID, NDRT, TOR, and TB), the sample characteristics (SAMPLE, AGE, and MALE) and the journal ranking (QUART), a PAM clustering, also known as partition around medoids or k-medoids clustering, with Gower distance was applied to the 150-observation database. The Gower distance is a common measure applied to clustering with mixed data [22], being calculated as the mean of partial dissimilarities among subjects and varying between 0 and 1. For numerical variables, the Gower distance between observations i and j for feature f can be computed as follows:

$$d_{ij}^{(f)} = \frac{|x_{if} - x_{jf}|}{R_f} \quad (1)$$

where x_i and x_j are the absolute difference between observations and R_f is the maximum observed range. For categorical variables, the Gower distance is equal to zero if observations i and j belong to the same category, and equal to 1 otherwise. Missing values are allowed, as the dissimilarities for a given feature are computed considering only the non-missing values.

PAM clustering requires to define a priori the number of clusters (k), thus iterations were run starting with $k=2$, using the data mining software R [35]. The selection of the number of clusters to retain was made according to both clustering performance and interpretability criteria. The silhouette coefficient (SC) [38]

was chosen to evaluate the clustering performance based on the pairwise differences between and within-cluster distances. The analysis of this index determined that the optimal number of clusters was 10, for which SC peaked at 0.496. However, for $k=10$, the clustering resulted in over-segmentation, isolating small groups that cannot be considered representative of general patterns. The first deceleration of the SC growth was observed between $k=4$ (SC=0.341) and $k=5$ (SC=0.359). Additionally, it was possible to obtain interpretable and meaningful results for $k=4$. For these reasons, four clusters were retained for this analysis.

Finally, to analyse the variations of takeover performance measures across the clusters, two one-way ANOVAs were conducted, considering TOT or CR as the dependent variable and the cluster coding as the independent variable. Post-hoc Tukey's HSD tests were conducted to analyse the significance of the differences between clusters. ANOVA was performed using the software IBM SPSS Statistics 26 [19].

4.3 Results and discussion

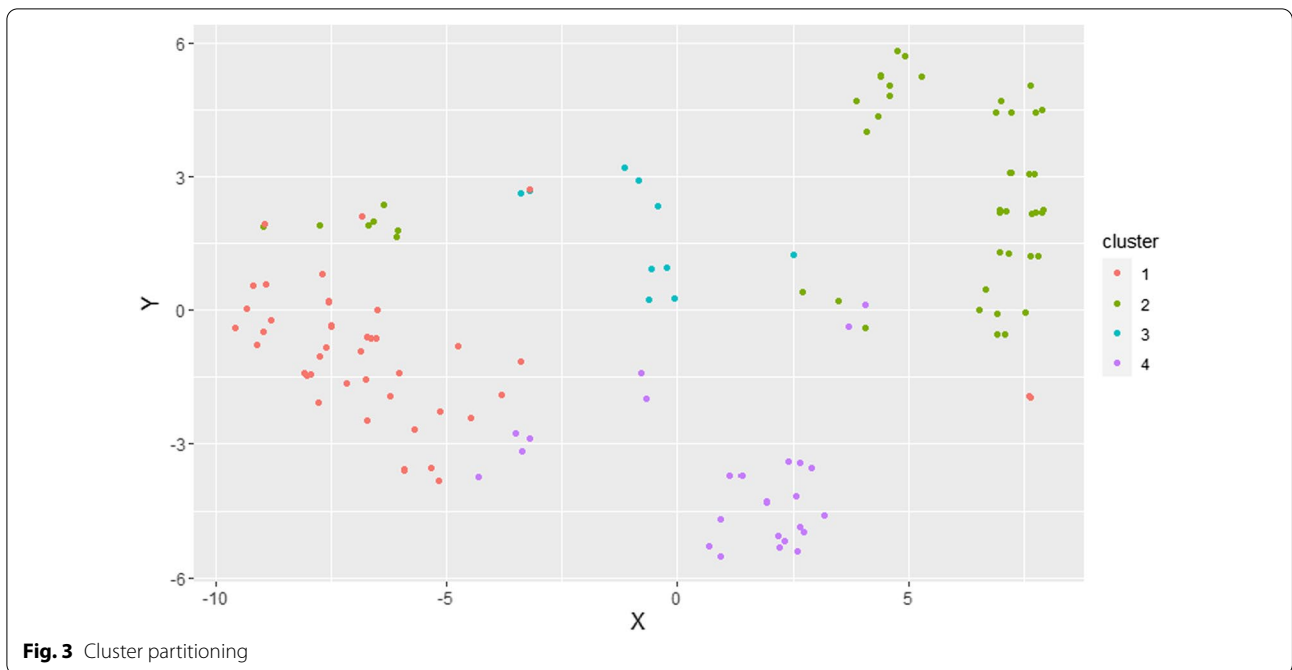
The centroids of the four groups resulting from the PAM clustering are presented in Table 5. The cluster partitioning is displayed on a two-dimensional space in Fig. 3. This chart was made using the *Rtsne* package in R [46, 47], which allows to construct a low dimensional embedding of high-dimensional data.

The relevance of the cluster partitioning is evidenced by the fact that each cluster's observations are grouped

Table 5 Cluster centroids

Variable	Cluster 1 (N = 47)	Cluster 2 (N = 57)	Cluster 3 (N = 15)	Cluster 4 (N = 31)
FID = low	0.04	0.11	0.60	0.10
FID = medium	0.21	0.00	0.27	0.90
FID = high	0.60	0.89	0.00	0.00
FID = n/a	0.15	0.00	0.13	0.00
NDRT = no	0.06	0.32	0.53	0.29
NDRT = optional	0.02	0.02	0.20	0.16
NDRT = mandatory	0.70	0.51	0.13	0.26
NDRT = n/a	0.21	0.16	0.13	0.29
TOR = no	0.02	0.21	0.73	0.00
TOR = yes	0.98	0.79	0.27	1.00
TB (s)	6.71	5.81	n/a	8.80
SAMPLE	33.44	29.47	24.15	34.97
AGE (years)	36.09	36.75	38.54	37.02
MALE (%)	51.39	52.91	56.21	60.19
QUART = Q1	0.96	0.12	0.13	0.13
QUART = Q2	0.00	0.88	0.07	0.81
QUART = Q3/Q4	0.04	0.00	0.80	0.06

N is the number of observations



approximately in the same zone of the chart of Fig. 3, with the exception, to a certain extent, of cluster 2. However, it should be noted that the data dimensionality reduction that enables this representation may cause that some clusters look less cohesive.

In relation to the cluster’s interpretation based on the obtained centroids, cluster 1 gathers observations with complex experimental settings, featuring NDRTs (72%), TORs (98%), and high-fidelity dynamic simulators (60%). Together with the fact that 96% of the observations were extracted from Q1 journal publications, this denotes that cluster 1 may be associated with studies featuring higher quality standards.

In contrast with cluster 1, the observations in cluster 3 are mainly associated with Q3 and Q4 journals (80% of observations), low-fidelity driving simulators (60%), and

no distractive task (53%) or TOR (73%). In combination with the lack of information on the time budget, those factors show a lower complexity of the studies represented in cluster 3.

Clusters 2 and 4 show some similarities, namely the prevalence of TORs and Q2 journals, as well as a more balanced distribution between the presence and the absence of NDRTs. The main differences are observed in the mean time budget, which is the lowest in cluster 2 (5.8 s) and the highest in cluster 4 (8.8 s), and in the type of driving simulator, with a prevalence of high-fidelity simulations (89%) in cluster 2 and medium-fidelity simulations (90%) in cluster 4. Therefore, both clusters seem to represent studies with intermediate levels of complexity, mostly differing in the simulation infrastructure and the adjustment of simulation parameters.

Table 6 Takeover time and crash rate descriptive statistics for each cluster

Cluster	TOT			CR		
	N	Mean (s)	Standard deviation (s)	N	Mean (%)	Standard deviation (%)
1	15	2.54	0.80	10	15.43	13.64
2	47	2.64	0.88	8	1.66	4.70
3	57	1.55	0.91	4	0.00	0.00
4	31	3.18	1.08	0	n/a	n/a
Total	150	2.61	0.99	22	7.62	11.86

The participants' characteristics did not present large variations across the four clusters: the mean age varied between 36 and 39 years old, the mean percentage of males between 51 and 60%, and the mean sample size between 24 and 35 participants. These small variations seem to confirm the use of relatively small convenience samples belonging to the same group (e.g., university students or workers at an R&D centre or car manufacturer).

The variations of takeover time and crash rate across clusters is depicted in Table 6. Crash rates are not available for any of the observations contained in cluster 4.

As determined by the one-way ANOVA, there are statistically significant differences between clusters, at the 1% level, in relation to the mean takeover time ($F(3,146)=10.988$, $p=0.000$) and mean crash rate ($F(2,19)=5.847$, $p=0.010$). A post-hoc Tukey's HSD test revealed that the takeover time was statistically significantly lower, at the 5% level, for cluster 3 in relation to clusters 1 ($p=0.020$), 2 ($p=0.000$) and 4 ($p=0.000$). The takeover time was also statistically significant higher for cluster 4 in relation to clusters 1 ($p=0.016$) and 2 ($p=0.041$). There was no statistically significant difference between clusters 1 and 2 ($p=0.956$). In relation to the crash rate, Tukey's HSD test showed that is statistically significantly higher in cluster 1 than in clusters 2 ($p=0.021$) and 3 ($p=0.039$). The mean crash rate is not statistically significantly different between clusters 2 and 3 ($p=0.959$).

The fact that cluster 3 presented the lowest mean takeover time and crash rate may be associated with the lower complexity of these studies. Despite that most experiments do not warn the driver for the need of manual intervention, NDRTs are notoriously absent or, at least, are optional, thus drivers may be more alert to potentially risky situations and to the urgent need of manual intervention. In practice, this type of experiments are useful as a baseline for comparison with more complex simulations rather than addressing the wide spectrum of driver-vehicle interactions occurring the real world.

Cluster 4 is associated with higher mean reaction times (crash rates are not available). Considering the higher mean time budgets and that all the observations presented a TOR, the higher reaction times may be explained by a timely anticipation of critical events and/or system limits (scheduled takeover), reducing the pressure on drivers to act quickly. This is the case of the studies by Dogan et al. [10], Vogelpohl et al. [49, 50] and Yoon and Ji [61], which have all their observations grouped in cluster 4.

Clusters 1 and 2 present similar mean takeover times (≈ 2.6 s), but the mean crash rate is significantly lower

in cluster 2. Both clusters are associated with high-fidelity simulators, the presence of TOR, NDRT engagement and similar mean time budgets. However, the prevalence of mandatory NDRT is higher in cluster 1 (70% versus 51%), and the absence of NDRT is much more notorious in cluster 2 (32% versus 6%), with comparable values for optional NDRT and missing data. Therefore, cluster 1 can be associated with more demanding experiments testing a larger array of risky behaviours leading to crashes. As previously mentioned, secondary task engagement has been widely associated with an increased crash risk [24, 53, 54, 62]. Although NDRTs have also been associated with higher reaction times [9, 58], some authors found contradictory or null effects [15, 17, 59]. The potential of NDRTs to increase drivers' situation awareness [3] narrows the consensus around the negative effects on takeover times and strengthens the importance of evaluating takeover performance using simultaneously time and quality measures [63].

The results may also denote some publication bias, with high-ranking journals favouring more complex studies in which it can be difficult to control for confounding effects. Clusters 1, 2 and 4 are widely dominated by Q1 or Q2 journal publications, while cluster 3, in which the lowest mean takeover time and crash rate are observed, is the only one characterized by a majority of Q3/Q4 publications. It should be remembered that cluster 3 is mainly associated with less complex experiments without TOR or NDRTs, but such studies are still relevant for the analysis of urgent takeover events and baseline monitoring conditions of the automated driving activity. As noted by Zhang et al. [65] regarding publication bias, "high and low takeover times could be regarded as equally interesting to authors, publishers, and editors."

As mentioned before, the study by Zhang et al. [65] can be seen as a reference in automated driving takeover reviews, particularly concerning a very detailed evaluation of the individual impacts of different variables on the mean takeover time. Despite the different objectives and approach, both Zhang et al. [65] and the present review achieved intuitive and consistent findings. Specifically, low urgency of manual control recovery, characterized by the presence of a TOR and long time budgets, seem to lead to higher takeover times. Zhang et al. [65] associated this with higher automation levels featuring increased capabilities of anticipating a TOR, in accordance to the SAE scale [39]. Both reviews concluded that avoiding NDRTs decreases takeover time. However, contrary to Zhang et al. [65], we did not find a relation between high-fidelity driving simulators and older samples of participants.

5 Conclusions

The transition to automated driving requires long and comprehensive research to address the risks introduced by new technology, particularly regarding the interaction between the automated vehicle, its driver, and the other road users. Because of the safety concerns related with an early maturity of automated systems and the lack of a widespread regulatory framework for real-world testing, driving simulators, together with closed testing facilities, have been used by the majority of researchers in the field.

This review systematizes the latest research performed in driving simulators focused on the manual control recovery in an automated vehicle, thus providing relevant information and guidance for the experimental design of future analyses. Takeover is a major safety concern associated with the intermediate stages of development of automated vehicles [11, 23], which has been demonstrated by an increasing number of studies published on this topic.

A descriptive analysis of 36 papers, selected according to different eligibility criteria, allowed understanding the most reproduced takeover situations and the most important measures used in the assessment of takeover performance. It was observed that researchers have been mainly dedicated to the analysis of critical situations that may reduce drivers' ability to reassume manual driving, such as distraction, drowsiness and passive fatigue. Different types of NDRTs, involving combinations of visual, auditory, motor and cognitive stimuli, have been tested against a wide array of more or less critical events. The takeover reaction time is, by far, the most used measure to assess takeover performance.

As expected, drivers' performance tends to decrease as the complexity of the situation increases, i.e., as the more immersive the NDRTs and the more critical the traffic situations are (e.g., play a game while an automation malfunction causes a lane departure). Nevertheless, some authors found that NDRTs may have, under certain conditions, a positive effect on drivers' situation awareness and reaction times [17, 59]. One of those situations is the takeover after a long period of automation [3], especially if timely anticipated by the system. In this case, secondary task engagement may help drivers to avoid boredom and drowsiness and to be more prepared to recover manual control.

The selected papers were also examined through a meta-analysis to identify different research patterns. Several variables characterizing the driving simulation settings, the participants' sample, and the takeover performance were extracted from the papers, together with the corresponding journal quartiles. A cluster analysis evidenced that more complex studies, i.e., featuring a wide array of simulation scenarios and developed in

high-fidelity dynamic driving simulators, are typically published in high-ranking journals (Q1 or Q2). These studies tend to reveal worse takeover performance indicators, as a result of the complexity of the analysed situations. In turn, simpler studies without secondary task engagement are more likely to be published in Q3 or Q4 journals. The lower takeover times and crash rates associated with these studies may not be representative of the many complex interactions between driver and vehicle in real-world driving. Notwithstanding, these studies allow for a better understanding of baseline conditions (e.g., no TOR and no NDRT) in comparison to more complex approaches.

This study has some limitations related to its conceptualization and methodology, but as any review, it also reflects the limitations of the current state-of-art. First, as the objectives were to provide a description of the main experimental conditions used for takeover research and to use a meta-analysis to establish patterns among the existing studies, the methodology can be classified as a between-study analysis that does not control for the confounding effects of within-study variables. For example, the results showed that higher takeover times are associated with the presence of TOR and higher time budgets, representing low-urgency events. Therefore, this does not mean that a sufficiently anticipated or scheduled TOR reduces takeover safety. In this sense, reducing takeover times should not necessarily be a priority for the advance of automated driving technology [65]. Designers and researchers should aim for the reduction of unexpected or sudden takeover events by increasing warning, communication and sensing capabilities. For a deep understanding of the individual effects of numerous variables on takeover time, please see Zhang et al. [65].

Second, the process of extracting variables from different sources may be a source of ambiguity, because it is not completely possible to assure that similar variables are measured in a similar way. An example of this issue is the instant considered to measure the reaction time after a TOR, which can be established according to different threshold values defining a human action on the steering wheel [28, 53] and/or pedals [26, 53, 62].

Third, as mentioned before, the crash rate is the most reported measure of takeover quality, but it is still absent from the majority of takeover performance assessment studies. This aspect is highlighted by the methodology adopted in this review, as one of the clusters is exclusively composed by observations with missing crash rates. A widespread availability of consistent takeover quality measures would allow for better insights on the crash risks associated with each experimental setup. Thus, for a deeper and more comprehensive assessment of takeover performance, future research should give more attention

to takeover quality indicators in addition to reaction times.

Fourth, it was observed that most of the selected studies used convenience samples with small variability of basic demographic indicators, such as gender and age. These characteristics did not vary significantly across the obtained clusters, which seems to confirm this issue. Future experimental research should then seek to increase the representativeness of the sample of participants in relation to the general population.

Finally, this review is limited to driving simulator studies, which represent the vast majority of takeover research conducted to date. Nevertheless, perceived crash risk is lower in virtual environment than in real-world driving, which could discourage fast reactions [65]. However, an increasing number of large-scale tests of automated vehicles on public roads, especially in the United States, together with open-source data initiatives by some OEMs and mobility companies, are expected to address this issue and boost R&D of automated driving technology in the next few years. Therefore, future reviews are also expected to address on-road conditions.

Abbreviations

ADAS: Advanced driver assistance systems; AGE: Participants' mean age; CR: Crash rate; DDT: Dynamic driving task; EC: European Commission; FID: Simulator fidelity; MALE: Percentage of male participants; NDRT: Non-driving-related task; QUART: Journal's highest quartile in Web of Science's 2020 Journal Citation Reports; SAMPLE: Participants' sample size; SC: Silhouette coefficient; SDLP: Standard deviation of lateral position; SuRT: Surrogate Reference Task; TB: Mean time budget; TOR: Takeover request; TOT: Mean takeover time; TRID: Transportation Research International Documentation; TTC: Time-to-collision; WoS: Web of science.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12544-021-00505-2>.

Additional file 1. Summary of the reviewed papers.

Acknowledgements

Not applicable.

Authors' contributions

Conceptualization: António Lobo and António Couto; Methodology: Sónia Soares, António Lobo, and António Couto; Formal analysis: Sónia Soares, António Lobo, and António Couto; Data curation: Sónia Soares and António Lobo; Writing – original draft preparation: Sónia Soares; Writing – review and editing: António Lobo, Sara Ferreira, Liliana Cunha, and António Couto; Visualization: Sónia Soares and António Lobo; Project administration: Sara Ferreira; Funding acquisition: Sara Ferreira, Liliana Cunha, and António Couto. Sónia Soares and António Lobo contributed equally to this work as first authors. All authors read and approved the final manuscript.

Funding

This work was financially supported by: Project PTDC/ECI-TRA/28526/2017—POCH-01-0145-FEDER-028526—funded by FEDER funds through COM-PETE2020—Programa Operacional Competitividade e Internacionalização (POCI) and by national funds (PIDDAC) through FCT/MCTES.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

Not applicable.

Author details

¹Research Centre for Territory, Transports and Environment, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal. ²Center for Psychology, Faculty of Psychology and Educational Sciences, University of Porto, Rua Alfredo Allen, s/n, 4200-135 Porto, Portugal.

Received: 3 April 2021 Accepted: 30 August 2021

Published online: 03 September 2021

References

- Alrefaie, M. T., Summerskill, S., & Jackson, T. W. (2019). In a heart beat: Using driver's physiological changes to determine the quality of a takeover in highly automated vehicles. *Accident Analysis & Prevention*, *131*, 180–190. <https://doi.org/10.1016/j.aap.2019.06.011>
- Blommer, M., Curry, R., Swaminathan, R., Tijerina, L., Talamonti, W., & Kochhar, D. (2017). Driver brake vs. steer response to sudden forward collision scenario in manual and automated driving modes. *Transportation Research Part F: Traffic Psychology and Behaviour*, *45*, 93–101. <https://doi.org/10.1016/j.trf.2016.11.006>
- Bourrelly, A., de Naurois, C., Zran, A., Rampillon, F., Vercher, J.-L., & Bourdin, C. (2019). Long automated driving phase affects take-over performance. *IET Intelligent Transport Systems*, *13*, 1249–1255. <https://doi.org/10.1049/iet-its.2019.0018>
- Choi, D., Sato, T., Ando, T., Abe, T., Akamatsu, M., & Kitazaki, S. (2020). Effects of cognitive and visual loads on driving performance after take-over request (TOR) in automated driving. *Applied Ergonomics*, *85*, 103074. <https://doi.org/10.1016/j.apergo.2020.103074>
- Dahlen, E. R., Martin, R. C., Ragan, K., & Kuhlman, M. M. (2005). Driving anger, sensation seeking, impulsiveness, and boredom proneness in the prediction of unsafe driving. *Accident Analysis & Prevention*, *37*(2), 341–348. <https://doi.org/10.1016/j.aap.2004.10.006>
- De Winter, J. C. F., Stanton, N., Price, J. S., & Mistry, H. (2016). The effects of driving with different levels of unreliable automation on self-reported workload and secondary task performance. *International Journal of Vehicle Design*, *70*, 297–324. <https://doi.org/10.1504/IJVD.2016.076736>
- DeGuzman, C. A., Hopkins, S. A., & Donmez, B. (2020). Driver takeover performance and monitoring behavior with driving automation at system-limit versus system-malfunction failures. *Transportation Research Record: Journal of the Transportation Research Board*, *2674*(4), 140–151. <https://doi.org/10.1177/0361198120912228>
- DinparastDjadid, A., Lee, J., Domeyer, J., Schwarz, C., Brown, T., & Gunaratne, P. (2021). Designing for the extremes: Modeling drivers' response time to take back control from automation using bayesian quantile regression. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *63*(3), 519–530. <https://doi.org/10.1177/0018720819893429>
- Dogan, E., Honnêt, V., Masfrand, S., & Guillaume, A. (2019). Effects of non-driving-related tasks on takeover performance in different takeover situations in conditionally automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *62*, 494–504. <https://doi.org/10.1016/j.trf.2019.02.010>
- Dogan, E., Rahal, M.-C., Deborne, R., Delhomme, P., Kemény, A., & Perrin, J. (2017). Transition of control in a partially automated vehicle: Effects of anticipation and non-driving-related task involvement. *Transportation*

- Research Part F: *Traffic Psychology and Behaviour*, 46, 205–215. <https://doi.org/10.1016/j.trf.2017.01.012>
11. Endsley, M. (2017). From here to autonomy: Lessons learned from human-automation research. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 59(1), 5–27. <https://doi.org/10.1177/0018720816681350>
 12. Eriksson, A., & Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 59(4), 689–705. <https://doi.org/10.1177/0018720816685832>
 13. European Commission. (2011). *White paper on transport. Roadmap to a single European transport area—Towards a competitive and resource efficient transport system*. Retrieved March 1, 2021, from https://ec.europa.eu/transport/themes/european-strategies/white-paper-2011_en
 14. European Road Safety Observatory. (2018). *Annual Accident Report 2018*. European Commission. https://ec.europa.eu/transport/road_safety/sites/default/files/pdf/statistics/dacota/asr2018.pdf. Accessed 1 March 2021.
 15. Feldhütter, A., Hecht, T., Kalb, L., & Bengler, K. (2019). Effect of prolonged periods of conditionally automated driving on the development of fatigue: With and without non-driving-related activities. *Cognition, Technology & Work*, 21, 33–40. <https://doi.org/10.1007/s10111-018-0524-9>
 16. Fisher, D. L., Rizzo, M., Caird, J. K., & Lee, J. D. (2011). *Handbook of driving simulation for engineering, medicine and psychology*. CRC Press.
 17. Gold, C., Happee, R., & Bengler, K. (2017). Modeling take-over performance in level 3 conditionally automated vehicles. *Accident Analysis & Prevention*, 116, 3–13. <https://doi.org/10.1016/j.aap.2017.11.009>
 18. Happee, R., Gold, C., Radlmayr, J., Hergeth, S., & Bengler, K. (2017). Take-over performance in evasive manoeuvres. *Accident Analysis & Prevention*, 106, 211–222. <https://doi.org/10.1016/j.aap.2017.04.017>
 19. IBM Corp. (2019). *Released. IBM SPSS Statistics for Windows, Version 26.0*. IBM Corp.
 20. International Organization for Standardization. (2012). *Road vehicles—Ergonomic aspects of transport information and control systems—Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems* (ISO/TS Standard No. 14198:2012). <https://www.iso.org/standard/54496.html>
 21. International Organization for Standardization. (2019). *Road vehicles—Ergonomic aspects of transport information and control systems—Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems* (ISO/TS Standard No. 14198:2019). <https://www.iso.org/standard/71509.html>
 22. Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. Wiley.
 23. Kyriakidis, M., De Winter, J. C. F., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., Martens, M. H., Bengler, K., Andersson, J., Merat, N., Reed, N., Flament, M., Hagenzieker, M., & Happee, R. (2019). A human factors perspective on automated driving. *Theoretical Issues in Ergonomics Science*, 20(3), 223–249. <https://doi.org/10.1080/1463922X.2017.1293187>
 24. Lin, Q., Li, S., Ma, X., & Lu, G. (2020). Understanding take-over performance of high crash risk drivers during conditionally automated driving. *Accident Analysis & Prevention*, 143, 105543. <https://doi.org/10.1016/j.aap.2020.105543>
 25. Litman, T. (2018). *Autonomous vehicle implementation predictions: Implications for transport planning*. Vitoria Transport Policy Institute. Retrieved March 1, 2021, from <https://www.vtpi.org/avip.pdf>
 26. Lodinger, N. R., & DeLucia, P. R. (2019). Does automated driving affect time-to-collision judgments? *Transportation Research Part F: Traffic Psychology and Behaviour*, 64, 25–37. <https://doi.org/10.1016/j.trf.2019.04.025>
 27. Louw, T., Kuo, J., Romano, R., Radhakrishnan, V., Lenné, M. G., & Merat, N. (2019). Engaging in NDRTs affects drivers' responses and glance patterns after silent automation failures. *Transportation Research Part F: Traffic Psychology and Behaviour*, 62, 870–882. <https://doi.org/10.1016/j.trf.2019.03.020>
 28. Louw, T., Markkula, G., Boer, E., Madigan, R., Carsten, O., & Merat, N. (2017). Coming back into the loop: Drivers' perceptual-motor performance in critical events after automated driving. *Accident Analysis & Prevention*, 108, 9–18. <https://doi.org/10.1016/j.aap.2017.08.011>
 29. Lu, Z., Zhang, B., Feldhütter, A., Happee, R., Martens, M., & De Winter, J. C. F. (2019). Beyond mere take-over requests: The effects of monitoring requests on driver attention, take-over performance, and acceptance. *Transportation Research Part F: Traffic Psychology and Behaviour*, 63, 22–37. <https://doi.org/10.1016/j.trf.2019.03.018>
 30. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed1000097>
 31. Naujoks, F., Höfling, S., Purucker, C., & Zeeb, K. (2018). From partial and high automation to manual driving: Relationship between non-driving related tasks, drowsiness and take-over performance. *Accident Analysis & Prevention*, 121, 28–42. <https://doi.org/10.1016/j.aap.2018.08.018>
 32. Payre, W., Cestac, J., Dang, N.-T., Vienne, F., & Delhomme, P. (2017). Impact of training and in-vehicle task performance on manual control recovery in an automated car. *Transportation Research Part F: Traffic Psychology and Behaviour*, 46, 216–227. <https://doi.org/10.1016/j.trf.2017.02.001>
 33. Payre, W., Cestac, J., & Delhomme, P. (2016). Fully automated driving: Impact of trust and practice on manual control recovery. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 58(2), 229–241. <https://doi.org/10.1177/0018720815612319>
 34. Prahl, A., & Enright, R. (2017). Forgive computers: The rise of automation and implications for counseling. *Counseling and Values*, 62, 144–158. <https://doi.org/10.1002/cvj.12056>
 35. R Core Team. (2018). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Retrieved February 15, 2021, from <http://www.R-project.org>
 36. Radlmayr, J., & Bengler, K. (2015). *Literaturanalyse und Methodenauswahl zur Gestaltung von Systemen zum hochautomatisierten Fahren* (FAT-Schriftenreihe 276). German Association of the Automotive Industry. Retrieved July 15, 2021, from <https://www.vda.de/dam/vda/publications/2015/fat-schriftenreihe-276.pdf>
 37. Rauffet, P., Botzer, A., Chauvin, C., Saïd, F., & Camille, T. (2020). The relationship between level of engagement in a non-driving task and driver response time when taking control of an automated vehicle. *Cognition, Technology & Work*, 22, 721–731. <https://doi.org/10.1007/s10111-019-00611-1>
 38. Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
 39. SAE International. (2018). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles* (SAE Standard No. J3016_201806). Retrieved March 1, 2021, from https://www.sae.org/standards/content/j3016_201806
 40. Scharfe, M. S. L., Zeeb, K., & Russwinkel, N. (2020). The impact of situational complexity and familiarity on takeover quality in uncritical highly automated driving scenarios. *Information*, 11(2), 115. <https://doi.org/10.3390/info11020115>
 41. Shen, S., & Neyens, D. (2017). Assessing drivers' response during automated driver support system failures with non-driving tasks. *Journal of Safety Research*, 61, 149–155. <https://doi.org/10.1016/j.jsr.2017.02.009>
 42. Simões, A., Cunha, L., Ferreira, S., Carvalhais, J., Tavares, J. P., Lobo, A., Couto, A., & Silva, D. (2020). The user and the automated driving: A state-of-the-art. In N. Stanton (Ed.), *Advances in human factors of transportation, advances in intelligent systems and computing* (vol. 964, pp. 190–201). Springer. https://doi.org/10.1007/978-3-030-20503-4_17
 43. Soares, S., Ferreira, S., & Couto, A. (2020). Driving simulator experiments to study drowsiness: A systematic review. *Traffic Injury Prevention*, 21(1), 29–37. <https://doi.org/10.1080/15389588.2019.1706088>
 44. Soares, S., Monteiro, T., Lobo, A., Couto, A., Cunha, L., & Ferreira, S. (2020). Analyzing driver drowsiness: From causes to effects. *Sustainability*, 12(5), 1971. <https://doi.org/10.3390/su12051971>
 45. Vacek, S., Schimmel, C., & Dillmann, R. (2007). Road-marking analysis for autonomous vehicle guidance. In *Proceedings of the 3rd European conference on mobile robots, Germany*.
 46. van der Maaten, L. (2015). Accelerating t-SNE using tree-based algorithms. *Journal of Machine Learning Research*, 15, 3221–3245. <https://doi.org/10.5555/2627435.2697068>
 47. van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.
 48. Vogelpohl, T., Gehlmann, F., & Vollrath, M. (2019). Task interruption and control recovery strategies after take-over requests emphasize need for measures of situation awareness. *Human Factors: The Journal of the*

- Human Factors and Ergonomics Society*, 62(7), 1190–1211. <https://doi.org/10.1177/0018720819866976>
49. Vogelpohl, T., Kühn, M., Hummel, T., & Vollrath, M. (2018). Asleep at the automated wheel—Sleepiness and fatigue during highly automated driving. *Accident Analysis & Prevention*, 126, 70–84. <https://doi.org/10.1016/j.aap.2018.03.013>
 50. Vogelpohl, T., Kühn, M., Hummel, T., Gehlert, T., & Vollrath, M. (2018). Transitioning to manual driving requires additional time after automation deactivation. *Transportation Research Part F: Traffic Psychology and Behaviour*, 55, 464–482. <https://doi.org/10.1016/j.trf.2018.03.019>
 51. Vogelpohl, T., Vollrath, M., Kühn, M., Hummel, T., & Gehlert, T. (2016). *Übergabe von hochautomatisiertem Fahren zu manueller Steuerung* (Forschungsbericht Nr. 39). German Insurers Accident Research, German Insurance Association. Retrieved July 15, 2021, from https://udv.de/sites/default/files/tx_udvpublications/fobe_39_hochautomfahren.pdf
 52. Walch, M., Mühl, K., Kraus, J., Stoll, T., Baumann, M., & Weber, M. (2017). From car-driver-handovers to cooperative interfaces: visions for driver-vehicle interaction in automated driving. In G. Meixner, & C. Müller (Eds.), *Automotive user interfaces, human-computer interaction series* (pp. 273–294). Springer. https://doi.org/10.1007/978-3-319-49448-7_10
 53. Wan, J., & Wu, C. (2018). The effects of lead time of take-over request and nondriving tasks on taking-over control of automated vehicles. *IEEE Transactions on Human-Machine Systems*, 48(6), 582–591. <https://doi.org/10.1109/THMS.2018.2844251>
 54. Wandtner, B., Schömig, N., & Schmidt, G. (2018). Effects of non-driving related task modalities on takeover performance in highly automated driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 60(6), 870–881. <https://doi.org/10.1177/0018720818768199>
 55. Wandtner, B., Schömig, N., & Schmidt, G. (2018). Secondary task engagement and disengagement in the context of highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 58, 253–263. <https://doi.org/10.1016/j.trf.2018.06.001>
 56. Williamson, A., Friswell, R., Olivier, J., & Grzebieta, R. (2014). Are drivers aware of sleepiness and increasing crash risk while driving? *Accident Analysis & Prevention*, 70, 225–234. <https://doi.org/10.1016/j.aap.2014.04.007>
 57. World Health Organization. (2018). *Global status report on road safety 2018*. Retrieved March 1, 2021, from <https://www.who.int/publications/i/item/9789241565684>
 58. Wu, C., Wu, H., Lyu, N., & Zheng, M. (2019). Take-over performance and safety analysis under different scenarios and secondary tasks in conditionally automated driving. *IEEE Access*, 7, 136924–136933. <https://doi.org/10.1109/ACCESS.2019.2914864>
 59. Wu, Y., Kihara, K., Hasegawa, K., Takeda, Y., Sato, T., Akamatsu, M., & Kitazaki, S. (2020). Age-related differences in effects of non-driving related tasks on takeover performance in automated driving. *Journal of Safety Research*, 72, 231–238. <https://doi.org/10.1016/j.jsr.2019.12.019>
 60. Wu, Y., Kihara, K., Takeda, Y., Sato, T., Akamatsu, M., & Kitazaki, S. (2019). Effects of scheduled manual driving on drowsiness and response to take over request: A simulator study towards understanding drivers in automated driving. *Accident Analysis & Prevention*, 124, 202–209. <https://doi.org/10.1016/j.aap.2019.01.013>
 61. Yoon, S. H., & Ji, Y. G. (2019). Non-driving-related tasks, workload, and takeover performance in highly automated driving contexts. *Transportation Research Part F: Traffic Psychology and Behaviour*, 60, 620–631. <https://doi.org/10.1016/j.trf.2018.11.015>
 62. Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accident Analysis & Prevention*, 78, 212–221. <https://doi.org/10.1016/j.aap.2015.02.023>
 63. Zeeb, K., Buchner, A., & Schrauf, M. (2016). Is take-over time all that matters? The impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident Analysis & Prevention*, 92, 230–239. <https://doi.org/10.1016/j.aap.2016.04.002>
 64. Zeeb, K., Haertel, M., Buchner, A., & Schrauf, M. (2017). Why is steering not the same as braking? The impact of non-driving related tasks on lateral and longitudinal driver interventions during conditionally automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 50, 65–79. <https://doi.org/10.1016/j.trf.2017.07.008>
 65. Zhang, B., De Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 64, 285–307. <https://doi.org/10.1016/j.trf.2019.04.020>
 66. Zhou, H., Itoh, M., & Kitazaki, S. (2020). Effect of instructing system limitations on the intervening behavior of drivers in partial driving automation. *Cognition, Technology & Work*, 22, 321–334. <https://doi.org/10.1007/s10111-019-00568-1>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)