Applied Ergonomics 112 (2023) 104057

Contents lists available at ScienceDirect

Applied Ergonomics

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

Training for the safe activation of Automated Vehicles matters: Revealing the benefits of online training to creating glaringly better mental models and behaviour

Siobhan E. Merriman *, Kirsten M.A. Revell, Katherine L. Plant

Human Factors Engineering, Transportation Research Group, Faculty of Engineering and Physical Sciences, Boldrewood Innovation Campus, University of Southampton, Burgess Road, Southampton, SO16 7QF, UK

ARTICLE INFO

Keywords:
Automated vehicles
Driver training
Mental models
Trust in automation
Workload
Mental demand

ABSTRACT

Automated Vehicle (AV) systems are expected to reduce the frequency and severity of on-road collisions. Unless drivers have an appropriate mental model for the capabilities and limitations of the automation, they may not activate the automation safely or appropriately on the road, potentially leading to a collision. As such, a training package (L4DTP) was developed to improve drivers' decisions and behaviour when activating an AV system and this was evaluated in a between-subjects simulator experiment. Drivers received no training (NT, control group), read an owner's manual (OM, experimental group 1: current training provision) or underwent the L4DTP (experimental group 2: new training programme). All drivers then experienced five scenarios in a driving simulator where they encountered road conditions which were safe and unsafe for activation. Their activation decisions, behaviour, trust in automation, workload and mental models were measured. This experiment found that drivers who read the OM or underwent the L4DTP made better activation decisions and showed better activation behaviour compared to drivers who received NT. Additionally, drivers who underwent the L4DTP found it easier, less demanding and felt under less time pressure when making their decisions, had to expend less effort to reach the same activation performance and had more appropriate and comprehensive mental models for when the automation can be activated compared to drivers who read the OM. This L4DTP can make roads safer by reducing collisions linked to poor activation decisions and behaviour. Therefore, there is the potential for a real benefit for society if this training programme is adopted into mandatory AV driver training.

1. Introduction

Over recent years, governments and vehicle manufacturers have invested considerable resources into the research and development of Automated Vehicle (AV) systems. They have the potential to improve sustainability by reducing the pollution and greenhouse gas emissions that are produced by vehicles (Greenblatt and Shaheen, 2015; Bagloee et al., 2016), improve traffic flow (Choi and Ji, 2015), improve mobility and convenience for those who are currently unable to drive (e.g. elderly, disabled: Choi and Ji, 2015) and improve road safety by reducing the frequency and severity of on-road collisions (Schoettle and Sivak, 2014).

The SAE (2018) define six levels of driving automation. These are summaried in Table 1.

Levels 1 and 2 AV systems are available to purchase on the vehicle market (e.g. Tesla's autopilot suite: National Transportation Safety Board, 2020), some vehicle manufactures are skipping the development of Level 3 AV systems because they are considered unsafe (e.g. Ford Motor Company, 2016; Volvo Cars, 2017), Level 5 AV systems are a long way off and some vehicle manufacturers planned to introduce Level 4 AV systems for personal use to the vehicle market by the mid-2020s (e.g. 2021 for Volvo and Ford (Ford Motor Company, 2016; Volvo Cars, 2017), 2024 for Daimler (2019) and 2025 for Honda (2017)). There is a timely need by manufacturers to focus on AV systems which are designed for personal use and possess some Level 4 capabilities, therefore this article will focus on this type of AV system (i.e. not purely passenger AV systems such as autonomous taxis or shuttles). More specifically, this article will focus on an AV system which can perform all driving tasks in highly reliable road conditions and can reach minimal risk conditions (i.e. has some Level 4 capabilities, see Table 1). However,

* Corresponding author.
E-mail addresses: S.Merriman@soton.ac.uk (S.E. Merriman), K.M.Revell@soton.ac.uk (K.M.A. Revell), K.Plant@soton.ac.uk (K.L. Plant).

https://doi.org/10.1016/j.apergo.2023.104057
Received 25 October 2022; Received in revised form 25 May 2023; Accepted 27 May 2023
Available online 6 June 2023

Table 1
The SAE's six levels of driving automation (SAE, 2018).

SAE Level	Name	Driver's Role(s)	Automation's Role(s)
0	No Automation	Performs all driving tasks	None
1	Driver Assistance	Performs all driving tasks	Provides support for steering or braking and acceleration in certain road conditions
2	Partial Automation	Monitors the vehicle and road environment and takes over control of the vehicle when required (e.g. system limitations or failures) or desired	Performs the steering, braking and acceleration of the vehicle in certain road conditions
3	Conditional Automation	No longer required to perform the monitoring task, but must take over control of the vehicle when required or desired	Performs all driving tasks (including monitoring) in certain road conditions
4	High Automation	No longer required to perform takeovers but can if desired	Performs all driving tasks (including monitoring) in certain road conditions and can transition to a minimal risk condition (e.g. safe stop, enter hard shoulder) during system failures or when limitations are reached
5	Full Automation	None	Performs all driving tasks in all driver-manageable road and environmental conditions

in moderately reliable road conditions, the AV system is not designed to perform as well. The AV system has enhanced Level 2 capabilities, so some monitoring of the road environment and the vehicle will be required (Merriman et al., 2023). This type of AV system was chosen in this article because this approach is being taken by some vehicle manufactuers (e.g. Audi, 2017). For clarity, the rest of this article will refer to this AV system as a Level 4 AV system.

Considerable research has investigated the impact of AV systems on driver behaviour. This has revealed that despite all the potential benefits (mentioned above), there are also many challenges with AV systems. One such challenge is ensuring drivers have an appropriate mental model for the capabilities and limitations of the automation (Merriman et al., 2021a). A mental model is a person's knowledge and understanding of the physical world, the behaviour of a system or the automation (Stanton and Young, 2005; Saffarian et al., 2012). Ensuring drivers have an appropriate mental model is important because research suggests that this can influence their trust in automation and behaviour when operating the automation. Although takeovers are not compulsory for the AV system that is being considered in this article (see above), most of the research has looked at the links between mental models and takeover behaviours. As such, research on takeovers is presented here. Research suggests that if drivers believe that the automation is more capable than it actually is, they may over-trust and over-rely on the automation, activate the automation in inappropriate road conditions (Lee and See, 2004; Korber et al., 2018) and not take over control of the vehicle when needed (Barg-Walkow and Rogers, 2016; Cahour and Forzy, 2009). However, if drivers believe that the automation is less capable than it actually is, they may distrust and under-rely on the automation and not use it when it is safe and appropriate to use (Lee and See, 2004; Koustanaï et al., 2012; Korber et al., 2018; Boelhouwer, et al., 2019). Therefore, drivers need to have an appropriate mental model so that they develop more appropriate trust levels and show more appropriate behaviours when operating the automation (e.g. better takeover behaviours: Hergeth et al., 2017; Korber et al., 2018; Sportillo et al., 2019).

Training is one intervention that has been used to develop and improve trainees' mental models (e.g. Hays et al., 1992; Marks et al., 2000). In the automation domain, some training programmes have been developed to improve drivers' mental models for AV systems (e.g. Beggiato and Krems, 2013; Boelhouwer et al., 2019; Ebnali et al., 2019; Manser et al., 2019; Sportillo et al., 2019; Krampell et al., 2020). However, these studies focussed on Levels 2 and 3 AV systems and only evaluated the effectiveness of the training programmes by measuring drivers' takeover knowledge, decisions and/or behaviour. This article focusses on an AV system which possesses some Level 4 capabilities (see above) and the links between drivers' mental models and their activation behaviour for the following reasons. Firstly, the AV system defined in this article is capable of reaching a minimal risk condition (see above), which means that drivers are not required to take over control of the vehicle. However, they are still in full control of activating the AV system, so ensuring drivers activate the automation safely on the road should have a greater focus in training programmes for this type of AV system compared to the takeover task. Secondly, although the SAE (2018) guidance suggests that systems with Level 4 capabilities will not operate unless all the required conditions are met (i.e. unless the road conditions are safe) and will reach minimal risk conditions when the limitations of the system are reached, the technology may not meet these minimal requirements, in a similar way to current Level 2 AV systems. For example, current Level 2 AV systems do not limit the use of the AV system to the conditions for which they were designed (National Transportation Safety Board, 2020). As the technology is new and road and weather conditions are highly variable and can change unpredictably, it is very unlikely that all the scenarios where the vehicle could be used, and that are needed to meet these requirements, will be tested. Therefore there is a risk that Level 4 AV systems are not always going to be able to detect safe and unsafe road conditions, which means that they may not reach minimal risk conditions or stop drivers from activating the automation in all unsafe road conditions. Consequently, it is important for drivers to have an appropriate mental model for the capabilities and limitations of Level 4 AV systems, as a backup for the activation and takeover tasks in case the automation does not respond appropriately. Finally, this AV system has Level 4 capabilities in some road conditions and enhanced Level 2 capabilities in other road conditions, therefore it is important for drivers to have an appropriate mental model for the capabilities and limitations of this AV system to help them decide whether the road conditions are safe or not and also whether monitoring is required or not.

Merriman, et al. (Under Review) took an initial step towards this need by developing an online video-based training programme to improve drivers' mental models for activating this type of AV system and evaluating its effectiveness against the current training method for AV systems (an owner's manual: OM) in a matched-pairs experiment. Both training programmes provided the same content to drivers; they described the road conditions which were safe and unsafe for activation and explained how the reliability of the automation would affect their activation tasks (monitoring for moderately reliable road conditions). However the online video-based training programme combined video-clips, questions and feedback to deliver this training to drivers whereas the OM listed the same information in a document for drivers to read. The evaluation found that the online training programme in combination with an OM led to a greater improvement in drivers' mental models for when the automation can be activated compared to an OM in isolation. However, the evaluation did not investigate whether drivers' improved mental models translated to their actual activation decisions and behaviour. In both the online training programme and evaluation, drivers watched video-clips and had an unlimited time to determine the reliability of the automation and decide whether it was safe and appropriate to activate the automation. However in reality, drivers have to retrieve the relevant information and make these decisions in quick time, in a dynamic road environment whilst manually controlling their vehicle on the road. On the road, there is a greater consequence if mistakes/errors are made. Drivers are under a greater time pressure and workload from the additional task of manual driving, and as time pressure and a high workload can have a negative effect on decision-making (Hwang, 1994; Zakay and Wooler, 1984), the training programme may be less effective in improving drivers' activation decisions and behaviour on the road. As such, simulator and on-road evaluations are needed to see whether this online training programme improves drivers' activation decisions and behaviour on the road.

Additionally, the online training programme only targeted one of the four tasks that are required to activate this type of AV system safely on the road. A recent training needs analysis conducted by Merriman et al. (2023) shows that drivers need to follow a four-step procedure to activate this type of AV system safely on the road (Cruise, 2018; DVSA, 2014; IAM RoadSmart, 2016; SAE, 2018; Stanton et al., 2021; Tesla, 2019). They need to:

1. Recognise the alerts,
2. Decide whether to activate the automation,
3. Position their vehicle appropriately and
4. Activate the automation.

The improved mental models will only help drivers decide whether to activate the automation or not (step 2). Therefore, the online training programme may need to be supplemented with further procedural training to teach drivers all tasks required to activate this AV system safely on the road.

One method that has been successfully used in the wider training literature to help trainees learn, remember and perform sequential procedural tasks and tasks/information where order is important are acronyms (Ehrman et al., 2003; Evans, 2007; Gibson, 2009; Kovar and Van Pelt, 1991). For example, in the driving domain, IAM RoadSmart use the acronym IPSGA to help drivers remember the tasks that need to be performed when approaching and dealing with hazards on the road (IAM RoadSmart, 2016). Similarly in the aviation domain, acronyms such as DODAR and DECIDE have been used to help pilots with their decision-making. The use of acronyms for AV systems is not new; researchers have used acronyms to help drivers remember handover and takeover procedures (e.g. HazLanFuSea: Stanton, et al., 2021; CHAT: Shaw et al., 2020).

An acronym is a mnemonic technique which uses the first letter from a list of target words (the words/information that needs to be remembered) to create a more memorable word (Evans, 2007; Gibson, 2009; McCabe et al., 2013; Putnam, 2015). This word can be real (e.g. FAST for the treatment of strokes) or made-up (e.g. IPSGA), however it must be pronounceable (Higbee, 2001; Radović and Manzey, 2019). Acronyms use guided principals to aid trainees' learning, understanding and memory. They help to chunk (group) information together which reduces the number of items that need to be remembered. They enhance encoding, retention and retrieval because the target information is broken down (organised) into something more meaningful, familiar, memorable and easier to remember (Higbee, 2001; Lewis et al., 2018; McCabe et al., 2013; Putnam, 2015). The whole word and each letter acts as a retrieval cue for the target information, making the information easier and quicker to learn, understand, remember and retrieve from memory (Higbee, 2001; Putnam, 2015; Lewis et al., 2018). Acronyms are particularly effective in helping trainees learn, remember and perform sequential procedural tasks as each letter represents the different task steps/actions that need to be performed and the order of the letters helps trainees remember the correct order of those steps (Radović and Manzey, 2019). As such, acronyms help trainees recall the desired actions in a timely manner.

The task of activating the AV system can be categorised as a sequential procedural task as drivers have to perform each sub-task in a set order (from 1 to 4, see above). As such, an acronym was created to help drivers learn and remember the four main tasks that are required to activate the AV system safely on the road. As the online video-based training programme from Merriman et al. (Under Review) helps drivers perform one of these tasks (decide), the online training programme was then embedded into (framed by) the acronym to create a training package (L4DTP), designed to help drivers activate the AV system safely on the road. To overcome the limitations with the previous evaluation (Merriman et al., Under Review), this article evaluates the effectiveness of this L4DTP in a more realistic setting by measuring drivers' activation decisions and behaviour in a driving simulator. This article will describe the development and evaluation of this new L4DTP.

1.1. Aim and design

The aim of this study was to evaluate the effectiveness of the new L4DTP when compared to no training (NT) and an OM. To achieve this, a between-subjects experiment was conducted. Drivers received NT, read an OM or underwent the L4DTP (online video-based training and acronym). All drivers then experienced five scenarios in a driving simulator where they encountered road conditions which were safe and unsafe for activation. Their activation decisions, behaviour, trust in automation, workload and mental models were measured. The following hypotheses were made:

H_1: The type of training that drivers receive will affect their decisions about whether it is safe and appropriate to activate the automation.
H_2: The type of training that drivers receive will affect their workload when making their decisions.
H_3: The type of training that drivers receive will affect the appropriateness of their trust in automation.
H_4: The type of training that drivers receive will affect the appropriateness and comprehensiveness of their mental models for when the automation can be activated.
H_{10}: The type of training that drivers receive will not affect their decisions about whether it is safe and appropriate to activate the automation.
H_{20}: The type of training that drivers receive will not affect their workload when making their decisions.
H_{30}: The type of training that drivers receive will not affect the appropriateness of their trust in automation.
H_{40}: The type of training that drivers receive will not affect the appropriateness and comprehensiveness of their mental models for when the automation can be activated.

2. Method

2.1. Participants

Forty-five drivers between the ages of 18 and 66 who held a valid and full UK driving licence were randomly allocated to one of the three training conditions (each group consisted of 15 drivers). Key demographics for each training group are displayed in Table 2. Chi-square tests showed that the number of males and females ($p = .714$) and total number of advanced drivers ($p = .146$) did not significantly differ between the three training groups. Independent *t*-tests showed that there were no significant differences between the three training groups in terms of mean age ($p = .503$), years of licence ($p = .125$), annual mileage ($p = .877$), internality score ($p = .536$), externality score ($p = .122$) and trust in automation ($p = .573$). Ethical approval was gained by the University's Faculty Ethics Committee (Ergo: 70097). Drivers received £20 as compensation for taking part.

2.2. Materials

2.2.1. Level 4 Automated Vehicle Driver Training Package

The L4DTP consisted of two components: online video-based training and an acronym. These are explained below.

Table 2
Key demographics in each training condition.

Demographic		NT		OM		L4DTP	
		N		N		N	
Gender	Males	9		9		9	
	Females	6		6		5	
	Non-Binary	0		0		1	
Advanced Drivers	Males	0		0		3	
	Females	0		1		0	
	Non-Binary	0		0		0	
	Total	0		1		3	
		M	SD	M	SD	M	SD
Age		29.80	12.89	35.87	16.01	31.80	13.93
Years of Licence		8.00	9.29	18.50	16.77	14.32	14.04
Annual Mileage		6114.29	4163.74	6750.00	3333.71	6750.00	3387.81
Internality Score		34.60	10.90	30.50	9.23	33.20	9.56
Externality Score		34.40	8.40	31.50	8.01	28.13	8.07
Trust in Automation		113.07	9.77	117.57	16.56	116.73	9.40

2.2.1.1. Online video-based training. This was the online video-based training programme that was developed and evaluated in Merriman et al. (Under Review). In summary, drivers first read information about the SAE Levels of Driving Automation and the road conditions where the AV system can and cannot function reliably. This information was based upon the Operational Design Domain (ODD) that was defined for the AV system in Merriman et al. (2023), after reviewing the SAE's (2018) report on the definitions and taxonomy of driving automation, ODDs for current AV systems (e.g. National Transportation Safety Board, 2020; Tesla, 2019; Toyota, 2019) and information from vehicle manufacturers about the development of their AV systems (e.g. Audi, 2017; Skoda, 2018). The ODD is summarised in Fig. 1. Drivers were told that it was safe to activate the automation in the highly and moderately reliable road conditions, however it was not safe to activate the automation in the highly unreliable road conditions. Drivers were also told that some monitoring would be required if the automation was only moderately reliable.

Then they were shown 20 video-clips of everyday driving scenes. Seventeen video-clips displayed road conditions where the AV system cannot function reliably (i.e. highly unreliable road conditions in Fig. 1), and three video-clips showed road conditions where the AV system can function reliably (i.e. highly and moderately reliable road conditions in Fig. 1). After viewing each video-clip drivers were asked three questions (left-hand side of Fig. 2). Once they had submitted their responses, they were given written feedback of the correct answer (e.g. right-hand side of Fig. 2). See Merriman et al. (Under Review) for more details.

2.2.1.2. RUDPA. Once drivers had completed the online video-based training, they were introduced to the acronym and were told that they must follow this procedure/sequence of steps to activate the automation safely on the road. This section describes the development of the acronym and the materials given to drivers.

2.2.1.2.1. Acronym development. The procedure detailed in Ullius (1997) and Lewis et al. (2018) was followed. The target information for the acronym was first identified. The training needs analysis conducted by Merriman et al. (2023) decomposed each automation-specific task (Activate the Automation, Human-Directed Control Transfer, Vehicle-Directed Control Transfer and Emergency Control Transfer) into four discrete sub-tasks/steps:

1) Drivers need to **understand** the state of the automation (e.g. listen to and read the alerts and understand that the automation can be activated, look at the road environment and understand that the limitations of the automation have been reached).
2) Drivers need to make a **decision** (e.g. decide whether to activate the automation or continue manual driving, decide whether to take over control of the vehicle or let the vehicle reach a minimal risk condition).
3) Drivers need to **prepare** for the chosen action (e.g. move into the left-hand lane so they can activate the automation, build a situation awareness of the road environment before taking over control of the vehicle).
4) Drivers need to **perform** the chosen action (e.g. activate the automation, deactivate the automation and take over control of the vehicle).

As these four steps underlie all the automation-specific tasks, the four steps were deemed as essential information for drivers to learn and remember and therefore worthy of an acronym. Although this training programme is focussing on the task of activating the AV system, future training programmes will be able to use this acronym to help drivers learn and remember the four main steps that are required to perform Vehicle- and Human-Directed Control Transfers (takeovers). As such, drivers will only need to learn and remember one acronym to operate this AV system safely on the road.

Next, key terms for each step were identified. Alternative terms and synonyms were identified to provide a larger choice of letters to use. Finally, the first letter(s) from each term were taken to create an acronym. Different letter and word combinations were tried until the chosen acronym was selected.

The acronym that was chosen for this training programme was **RUDPA** (**R**ecognise, **U**nderstand, **D**ecide, **P**repare and **A**ct) as drivers

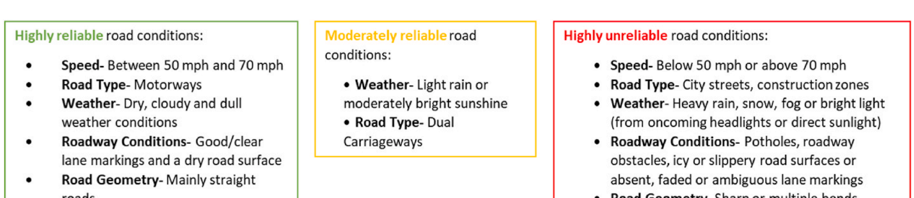


Fig. 1. The highly reliable, moderately reliable (safe) and unreliable (unsafe) road conditions for the AV system.

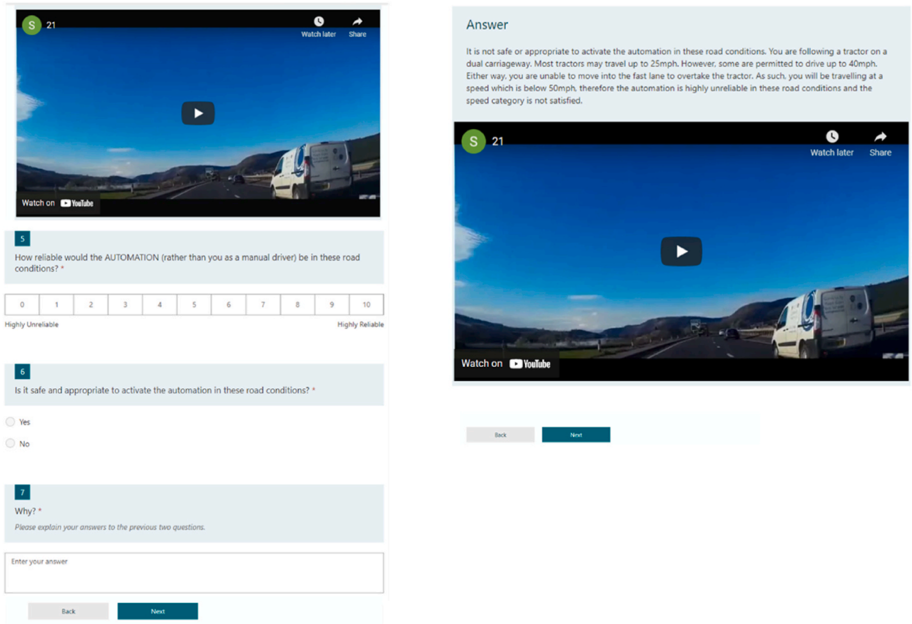


Fig. 2. An example video-clip, questions (left) and written feedback (right) in the online video-based training component of the L4DTP (see Merriman et al. (Under Review) for more details). SAE (2018) suggests that Level 4 AV systems may be able to detect and/or predict some unsafe road conditions. However to ensure drivers developed an accurate mental model for the AV system, drivers were required to make all the decisions.

have to perform all five steps/tasks (letters) in sequence to successfully and safely activate the automation. As this acronym can also be used to perform Human- and Vehicle-Directed Control Transfers, drivers will have a clear and standardised operating procedure to follow to operate this AV system safely on the road (Casner and Hutchins, 2019). In the aviation domain, standard operating procedures are used to ensure pilots use autopilot systems appropriately and perform appropriate behaviours at the right time (Degani and Wiener, 1997; Norman, 1988). Therefore, as drivers will be able to use this acronym to perform all tasks required to safely operate the AV system, frequent use will make this desired procedure (and associated tasks and behaviours) become automatic.

Additionally, this acronym satisfies the requirements for a successful and effective acronym: high discriminability (each letter-word pair should have a unique meaning) and associability (there should be an association/direct link between the acronym and the target information) (Bellezza, 1981; Kovar and Van Pelt, 1991; Ullius, 1997; Higbee, 2001). Each letter-word pair relates to a different step/task that needs to be performed and each word directly relates to the stage that it is representing (e.g. recognise for the recognising stage). This makes the words and associated behaviours/tasks easy to understand and remember.

2.2.1.2.2. Acronym materials. Once drivers had completed the online video-based training, they were presented with the RUDPA acronym. The task of activating the AV system was framed by the RUDPA acronym to demonstrate how this acronym should be used to activate the automation safely on the road. To do this, the sub-tasks and operations for the task of activating the automation from the training needs analysis conducted by Merriman et al. (2023) were used to decompose each RUDPA stage into separate tasks and behaviours and the training programme explained the roles, responsibilities, tasks and behaviours that drivers needed to perform at each stage. For example, for the stages “R” and “U”, drivers were told to listen to and read the auditory and visual alerts as these will tell them that the automation can be activated (Fig. 3).

2.2.2. Owner’s Manual

The nine-page OM that was developed in Merriman et al. (Under Review) was used in this evaluation. The OM described the systems that were present, the sensors and cameras used, the capabilities and limitations of each system, the road conditions which effect the reliability of the automation and when takeover requests may occur, the automation alerts and the activation controls (e.g. Fig. 4).

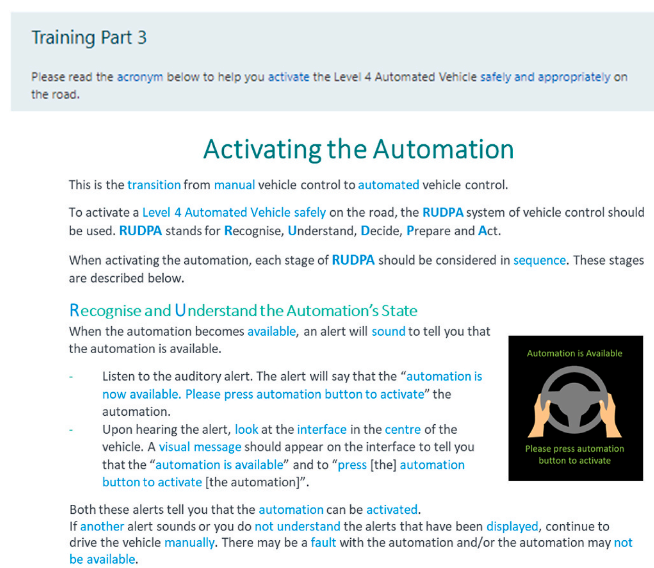
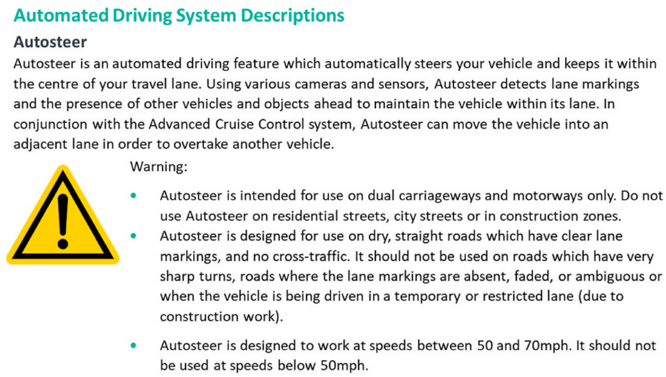


Fig. 3. Each RUDPA stage was decomposed into the tasks and behaviours that drivers needed to perform at that stage.

2.2.3. Training programme comparison

Both training programmes (L4DTP and OM) provided drivers with the same information (content), however the L4DTP positioned/framed this information in a step-by-step procedure (RUDPA) and used active (online video-based training) and passive (written information) methods to train drivers, whereas the OM did not use framing and only used passive training methods to train drivers. For example, for the task step “Recognise the automation is available alert”, the OM displayed images of the visual alerts, described the auditory alerts and explained in text


Automated Driving System Descriptions

Autosteer

Autosteer is an automated driving feature which automatically steers your vehicle and keeps it within the centre of your travel lane. Using various cameras and sensors, Autosteer detects lane markings and the presence of other vehicles and objects ahead to maintain the vehicle within its lane. In conjunction with the Advanced Cruise Control system, Autosteer can move the vehicle into an adjacent lane in order to overtake another vehicle.

Warning:

- Autosteer is intended for use on dual carriageways and motorways only. Do not use Autosteer on residential streets, city streets or in construction zones.
- Autosteer is designed for use on dry, straight roads which have clear lane markings, and no cross-traffic. It should not be used on roads which have very sharp turns, roads where the lane markings are absent, faded, or ambiguous or when the vehicle is being driven in a temporary or restricted lane (due to construction work).
- Autosteer is designed to work at speeds between 50 and 70mph. It should not be used at speeds below 50mph.

Fig. 4. An excerpt from the OM presented to the OM group during the evaluation of the L4DTP.

when they will occur. The L4DTP provided drivers with the same information, but it was framed under R and U in RUDPA (see Fig. 3). For the task step "Determine whether it is safe and appropriate to activate the automation", the OM displayed this information as a list of warnings (e.g. Fig. 4) whereas the L4DTP provided this information in online video-based training (e.g. Fig. 2) and framed it under D in RUDPA.

2.2.4. Questionnaires

Drivers completed seven questionnaires during this study. The first was a demographics questionnaire which asked about the driver's age, gender, annual mileage, years of licence, advanced driving qualifications, familiarity with Advanced Driver Assistance Systems and current training for AVs. The second was Montag and Comrey's (1987) 30-item Driving Internality and Driving Externality questionnaire. Drivers rated the extent to which they agreed with the 30 statements on a six-point Likert scale, ranging from 0-Disagree very much to 5-Agree very much. This questionnaire measured drivers' locus of control and provided one score for internality and one score for externality. The third questionnaire was the Total Trust in Automation questionnaire (Gold et al., 2015). This is a 35-item questionnaire which measured drivers' trust according to six subscales: Discharge of the driver due to the automation, Safety gains, Safety hazards, Trust in automation, Perceived control of conduct and Intention to use. Drivers rated the extent to which they agreed with the statements on a five-point Likert scale (from 1-Strongly disagree to 5-Strongly agree). These three questionnaires were completed online at the start of the study. The fourth questionnaire comprised the "Yes/No" and "Why" questions from the online training component of the L4DTP (see left-hand side of Fig. 2). All drivers completed these questions five times after they experienced each simulator scenario, to confirm their decision about whether they thought it was safe and appropriate to activate the automation (or not) and to gain an understanding of their reasoning for their decision. The drivers' reasoning was subsequently used to measure their mental models (see section 2.4). At the end of the study, drivers completed three questionnaires. The first was the Total Trust in Automation questionnaire (Gold et al., 2015). The second was Hart and Staveland's (1988) NASA-TLX which measured drivers' workload according to six subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration. Drivers were required to complete ratings on these dimensions from very high to very low (or from perfect to failure for performance) on 21-point scales. The final questionnaire asked drivers about their experiences in the simulator task and the training that they would find helpful to understand and remember the different tasks involved in activating the automation (open response).

2.2.5. Simulator tasks

To evaluate the effectiveness of the L4DTP, one terrain and five scenarios were created in a driving simulator. These are explained below.

2.2.5.1. Simulator description. The fixed-based driving simulator at the University of Southampton was used (Fig. 5). The simulator comprised a static Land Rover Discovery Sport's vehicle and three over-head forward projectors which provided a 135° field of view. An additional over-head rear projector provided a rear-view image of the driving scene, which could be viewed through the rear-view mirror. The door mirrors and dashboard were simulated using mini-LCD monitors. All primary (steering, throttle, brake) and secondary vehicle controls (buttons and switches) were fully functional. Sound was sent from the simulator software to the vehicle's audio system. All scenarios were created using AVSimulation's SCANeR software (version 1.8).

2.2.5.2. Chosen scenarios. The ODD presented in Fig. 1 defines numerous road conditions where it is unsafe to activate the AV system (i.e. the highly unreliable road conditions in Fig. 1). As such, many different scenarios could be created to evaluate the effectiveness of the L4DTP. To reduce the number of potential scenarios, attention was directed to the scenarios which drivers found particularly difficult in the previous evaluation (Merriman et al., Under Review). This was defined as the scenarios where the majority of drivers made mistakes before undergoing training (at test one), and/or the scenarios which showed mixed results after drivers underwent training (test two, i.e. the scenarios where some drivers still believed that it was safe to activate the automation after undergoing training). This resulted in five scenarios: poor lane markings, sharp bends, potholes, glare and debris/obstacles on the roadway. Between 72.9% and 96.9% of drivers incorrectly believed that it was safe to activate the automation in these scenarios before training, and between 35.4% and 46.9% continued to make these mistakes after training. However, issues with the simulator software meant that the poor lane markings and sharp bends scenarios could not be created. The remaining three scenarios were developed. Additionally two scenarios were added to represent road conditions which were safe and appropriate for activation (i.e. the highly and moderately reliable road conditions in Fig. 1). The development of these five scenarios is explained below.

2.2.5.3. Terrain design. A motorway terrain was created. The road had three lanes in the same direction, a hard shoulder and a central reservation. The speed limit was adjusted to 70 mph (112.65 km/h) and the vehicle type was restricted to highway vehicles (no pedestrians or cyclists). Fields, woodland, warehouses, buildings and cattle were added to the surrounding landscape.

2.2.5.4. Scenarios. The five scenarios were then developed on this terrain. The creation of these scenarios is described below.

2.2.5.4.1. Scenario 1: Appropriate Conditions 1. The target vehicle (the vehicle that drivers would be driving) was placed in the left-hand

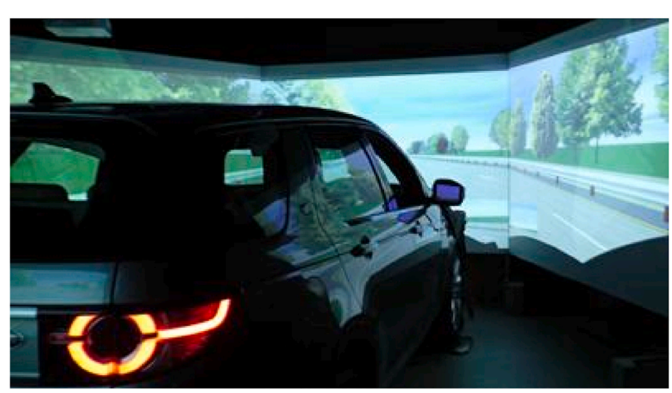


Fig. 5. The driving simulator that was used in the evaluation of the L4DTP.

lane of the motorway. Cars, vans, lorries and motorbikes were added in front and behind the vehicle, across all three lanes. The traffic in the left-hand lane was given a maximum speed restriction of 70 mph (112.65 km/h). To simulate natural overtaking traffic, the traffic in the centre lane was given a maximum speed restriction of 72 mph (96.56 km/h) and the traffic in the right-hand lane was given a maximum speed restriction of 75 mph (120.70 km/h).

To simulate safe/appropriate road conditions, this scenario was based upon the highly reliable road conditions described in Fig. 1. The scenario was in a motorway setting (see terrain above), lane markings were present and clear throughout, the road surface was dry, the road was mainly straight (no sharp or multiple bends), the weather was dry, cloudy and dull and the target vehicle was allowed to travel at a speed of up to 70 mph (no traffic, hazards or road conditions prevented drivers from reaching a speed of 70 mph, see Fig. 6).

2.2.5.4.2. Scenario 2: Appropriate Conditions 2. This scenario replicated the characteristics of the "Appropriate Conditions 1" scenario, however the landscape and traffic volume were modified. The first scenario represented a motorway in the countryside, as there were limited buildings but plenty of trees and animals surrounding the roadway. The second appropriate scenario represented a motorway in a built-up area, as there were less trees and animals but more buildings in the surrounding landscape (Fig. 7). Additionally, more traffic was added to the "Appropriate Conditions 2" scenario. The landscape and traffic volume were manipulated because in the previous evaluation (Merriman et al., Under Review), these variables influenced drivers' decisions about whether it was safe and appropriate to activate the automation. Before undergoing training, two common reasons why drivers believed that it was unsafe to activate the automation were "busy roads" or "lots of traffic" (vs "clear road", "lack of traffic") and "urban" (vs "rural roads"). This suggests that drivers were more uncertain about activating the automation in high traffic volumes (even if the traffic is steady) and in built-up areas, even though these variables are not highlighted as inappropriate road conditions (see Fig. 1). Therefore, these variables were manipulated in these scenarios.

These two scenarios formed the basis of the remaining three scenarios.

2.2.5.4.3. Scenario 3: Potholes. The "Appropriate Conditions 1" scenario was modified to add potholes, cracks and road deformations on the road surface at random locations during the two-minute decision phase. Road deformations were added 5, 15, 30, 45, 60, 73, 80, 90, 100 and 110 seconds after the automation became available. Additionally, to maintain consistency with the glare scenario where the torch was on throughout the whole scenario (see below) and to prevent drivers from making a good assessment about the road before the automation became available, small cracks were also added 10 and 20 seconds before the automation became available. Fig. 8 displays some of the road deformations that were used. In some locations, small cracks were added (e.g. top-left image) and in other locations big deformations were added (e.g. bottom-left image). The steering-wheel vibrated and turned when drivers drove over the road deformations.

2.2.5.4.4. Scenario 4: Debris/Obstacles. The "Appropriate Conditions 2" scenario was modified to add obstacles and debris in the left-hand lane at random locations during the two-minute decision phase. Tyres, traffic cones, barrels and gravel piles were placed on the road surface 9, 30, 56, 75, 85 and 103 seconds after the automation became available (Fig. 9). Some obstacles (tyres, traffic cones and barrels) required drivers to change lanes to avoid them. As such, some traffic was removed from the centre lane to ensure drivers had enough space to safely change lanes when required. Additionally, triggers were added to direct the traffic in the left-hand lane to change lanes to avoid the obstacles on the road. In each case, triggers were added six seconds before each vehicle hit the obstacle.

2.2.5.4.5. Scenario 5: Glare. The "Appropriate Conditions 1" scenario was used, however a torch was placed on the roof of the Land Rover and directed towards the front screen to reduce visibility and create glare (Fig. 10).

2.3. Procedure

Drivers were randomly allocated to one of the three training conditions. All drivers first completed the demographics questionnaire, locus of control questionnaire and trust in automation questionnaire online using Microsoft Forms. The NT group then went straight to the driving simulator (top level in Fig. 11). The only information this group were given about the AV system was a verbal description of the automation's capabilities (see introduction). In contrast, the other two groups completed their allocated training programme. The L4DTP group completed the online video-based training from Merriman et al. (Under Review) and read about the acronym (RUDPA) that they should use to activate the automation safely on the road (see section 2.2.1, bottom level in Fig. 11). The OM group read the OM (middle level in Fig. 11). To ensure all drivers received the same amount of time for training, the OM group were given 45 minutes to read the OM. This time was based upon the average amount of time that it took drivers to complete the L4DTP in pilot testing. Most drivers read the OM in approximately 15 minutes. When drivers finished early, they were reminded of the time that they had left and were told to revisit the materials. However if they were ready to move on, they were taken to the next task. Once they had completed their allocated training programme, they were offered a five-minute comfort and screen break.

Then all drivers underwent a practice drive in the driving simulator to get used to the vehicle and controls. They were shown the visual and auditory alerts on the central vehicular interface (see Fig. 12), practiced activating the automation, were informed that in all scenarios the automation will always become available and were told their goal and tasks (to activate the automation whenever it is safe and appropriate to do so and to always be in the left-hand lane when making their decisions).

They then experienced the five scenarios in the driving simulator. All

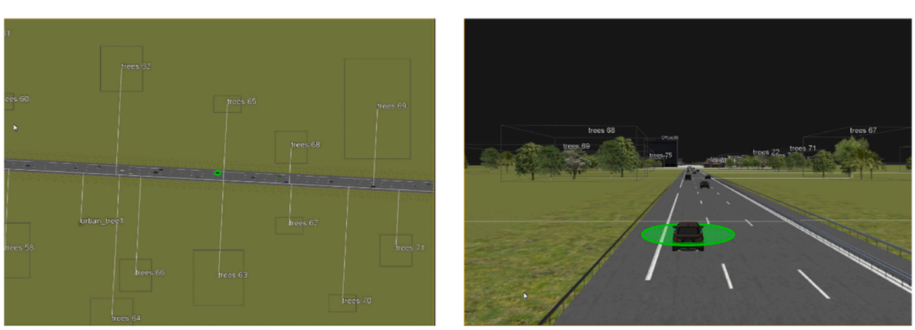


Fig. 6. The first appropriate road conditions scenario used in the evaluation of the L4DTP. The road was straight, line markings were clear and the weather was cloudy and dry.

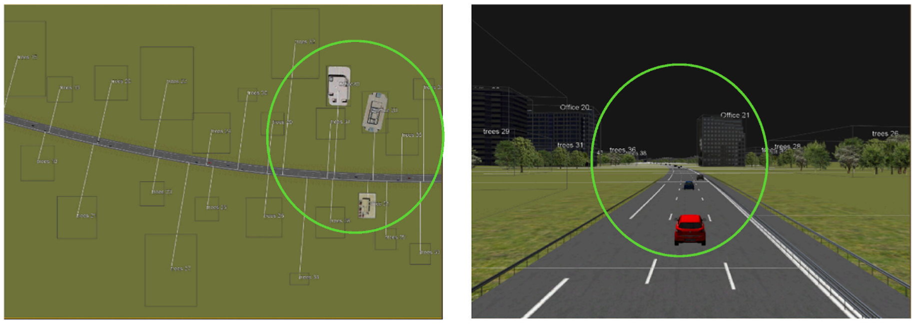


Fig. 7. The second appropriate road conditions scenario used in the evaluation of the L4DTP. More traffic and buildings were added to this scenario to simulate an urban road environment and a busier road.

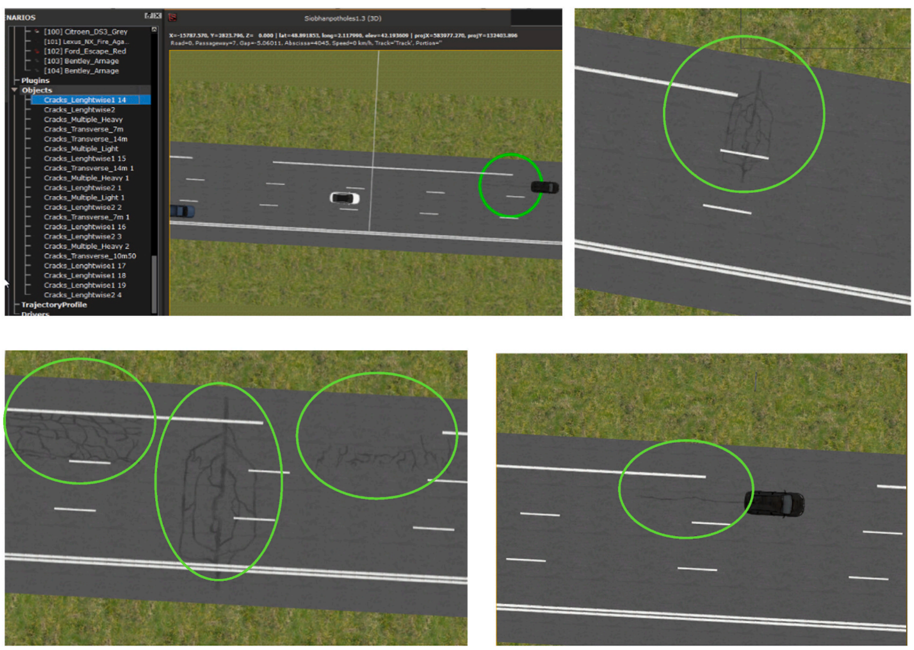


Fig. 8. The potholes scenario used in the evaluation of the L4DTP. Example potholes have been circled.

Fig. 9. Example obstacles used in the debris scenario to evaluate the L4DTP: a gravel pile (left), barrel (middle) and traffic cone (right).

scenarios followed the same format (Fig. 12). Drivers began the scenario driving the vehicle manually. After 30 seconds, they received an alert that the automation is available. They were then given two minutes to decide whether it was safe and appropriate to activate the automation, although they could make their decision at any point during this time. The two minutes were counted down on the central vehicular interface (Fig. 13). During this time, they were either exposed to road conditions which were safe and appropriate for activation (two appropriate scenarios) or road conditions which were unsafe and inappropriate for activation (debris/obstacles, potholes and glare scenarios). Present day technology means that the AV system may be able to detect and/or predict some of these unsafe road conditions (in which case the automation would not become available, and the alerts would not sound/appear). However for the purpose of this study and to ensure drivers developed an accurate mental model for when the automation can be activated, the automation always became available, and drivers were required to make all the decisions.

If they decided that it was safe and appropriate to activate the automation, they were told to activate the automation (i.e. go through the whole of RUDPA) and to press the hazard lights. If however they decided that it was unsafe and inappropriate to activate the automation, they were told to continue manual driving (i.e. stop at D in RUDPA) and to press the hazard lights. The hazards lights were a cue to the researcher that the driver had made a decision and so the scenario could be terminated. The scenario was terminated as soon as the hazard lights were pressed. If drivers had not made a decision after the two minutes,

Fig. 10. The glare scenario used in the evaluation of the L4DTP, showing the light source (left), outside view (middle) and driver view (right).

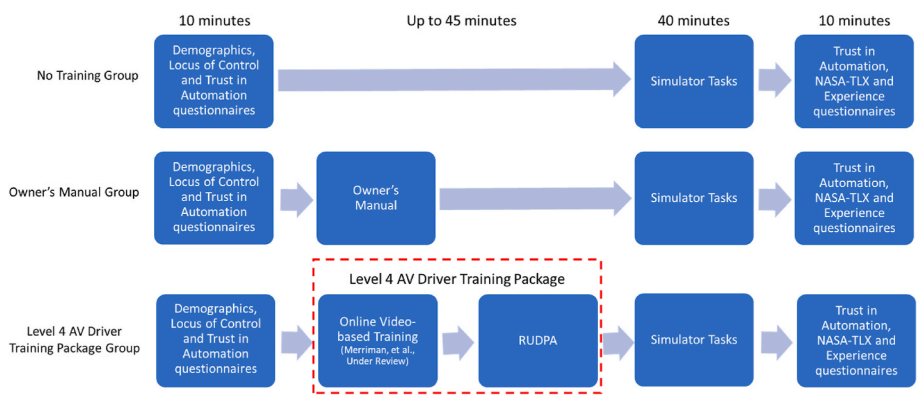


Fig. 11. Procedure for the three training conditions. The training time varied for each group. It was approximately 5 minutes for the no training group, 15 minutes for the OM group and 45 minutes for the L4DTP group.

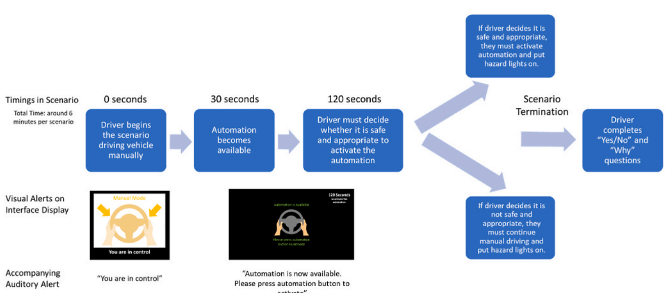


Fig. 12. Sequence for all five scenarios with associated timings and visual and auditory alerts.

the researcher instructed them to make an immediate decision. Once a decision had been made and they performed the appropriate behaviour, the scenario was terminated. No driver failed to make a decision after this prompt. At the end of each scenario, drivers completed the "Yes/No" and "Why" questions whilst seated in the simulator. Then the next scenario was played. To reduce order effects, the order of the scenarios was randomised for each driver using a random number generator. Once all five scenarios had been completed, drivers answered the final three questionnaires (trust in automation, NASA-TLX and experience questionnaires) and were instructed to base their answers on the simulator task that they had just completed.

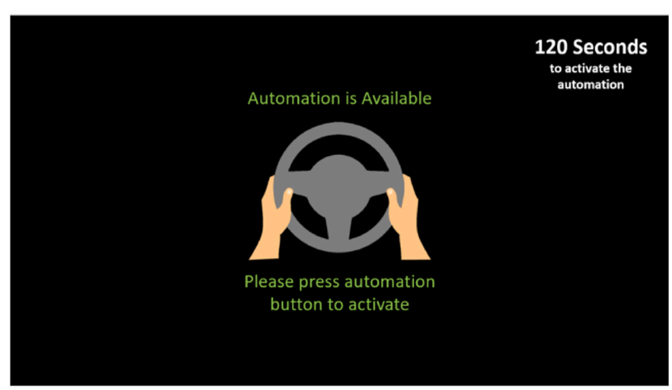


Fig. 13. Countdown timer on the central vehicular interface to tell drivers how much time they had before a decision was required.

2.4. Study design and analysis

To analyse the effectiveness of the L4DTP, a between-subjects experimental design was employed. The independent variable was the type of training programme. This was a between-subjects variable which had three levels: NT (control group), OM (experimental group 1: current training provision) and L4DTP (experimental group 2: new training programme).

There were five dependent variables, four quantitative and one qualitative:

1) Drivers' decisions about whether it was safe and appropriate to activate the automation in the five simulator scenarios. For each scenario, one point was awarded for the correct decision (i.e. drivers activate the automation in the safe/appropriate scenarios and do not activate the automation in the unsafe/inappropriate scenarios), creating a maximum score of five points.
2) Frequency and type of successes and errors on the simulator task, using the principles of signal detection theory (Green and Swets, 1966). In this approach, a signal detection grid is used to capture the frequency of hits, misses, false alarms and correct rejections when drivers' decisions are compared to the correct decisions on the simulator task. In this evaluation, the categories were defined as follows:
   - Hit: It was safe and appropriate to activate the automation and drivers decided to activate it.
   - Miss: It was safe and appropriate to activate the automation, but drivers decided to not activate it.
   - False Alarm: It was not safe and appropriate to activate the automation, but drivers decided to activate it.
   - Correct Rejection: It was not safe and appropriate to activate the automation and drivers decided to not activate it.

Therefore, there were two successes (hits and correct rejections) and two errors (misses and false alarms). Across the five simulator scenarios, drivers could score a maximum of two hits or two misses (on the two appropriate scenarios) and a maximum of three false alarms or three correct rejections (on the potholes, debris and glare scenarios). The number of hits, misses, false alarms and correct rejections for each participant were counted, following the guidance from Stanton et al. (2022).

3) Change in drivers' trust in automation from before training to after training. This was calculated by measuring the difference between drivers' before and after training trust scores (after training trust score minus before training trust score).
4) Drivers' workload on the simulator task. Drivers' scores on the six workload subscales and their total workload (summation of all six scores) was calculated.

These four dependent variables were analysed quantitatively. Where variables met the assumptions for parametric testing, a one-way independent measures ANOVA was used, and significant effects were broken down using the Bonferroni post-hoc test. If however a variable did not meet the assumptions for parametric testing, a Kruskal-Wallis Test was performed, and significant effects were broken down using Mann-Whitney U Tests, with the Bonferroni corrected alpha level of $p = .017$.

The fifth dependent variable was the appropriateness and comprehensiveness of drivers' mental models for when the automation can be activated. A variety of methods have been used to measure mental models (Rowe and Cooke, 1995). Traditionally researchers have measured mental models by looking at the manifestation of them in participants' performance, behaviour or decision-making on a related task (Gentner and Stevens, 1983). This is based upon the assumption that if participants have been taught the correct mental model, they will use that correct mental model in the task and this will be demonstrated in their behaviour (i.e. correct behaviour means correct mental model). However, these studies did not get participants to confirm whether they used the correct mental model or not (e.g. through a questionnaire, verbal protocol), leading to questions about whether the training had actually changed/improved participants' mental models and whether these improved mental models or a different mechanism had improved their performance. In the AV driver training literature, questionnaires have been used to measure drivers' mental models. For example, Blömacher et al. (2020), Beggiato et al. (2015) and Forster et al. (2019) provided statements on the AV system's functionality and asked drivers to provide ratings of agreement or correctness using scales. Boelhouwer et al. (2019) measured drivers' mental models by recording their decisions and their reasoning for their decisions in takeover scenarios. Therefore, to overcome the limitation with traditional mental model research, this same approach was used in this study to measure the appropriateness and comprehensiveness of drivers' mental models for when the automation can be activated.

A deductive thematic analysis was performed on the reasoning that drivers gave for their decisions on the simulator task ("why" questions in the simulator questionnaire, see section 2.2.4), using the guidelines from Braun and Clarke (2006). The drivers' reasoning for their decisions in the five simulator scenarios were read and road conditions relating to the five environmental categories defined in Fig. 1 (speed, road type, weather, roadway conditions, road geometry) were highlighted. For example, words and phrases related to the weather such as rain, wet, clear day, fog, glare, dry were coded under the weather category. Similarly, words and phrases related to bends, curves, hills or the straightness of the road were coded under the road geometry category. Road conditions which did not fit into one of the five categories were coded under a sixth "other" category. Then, the coded words for each category were reviewed to ensure they fit with the overall category.

3. Results

3.1. Decisions

There was a significant difference in the total number of correct decisions that drivers made on the simulator task between the three training conditions ($H(2) = 14.41$, $p = .001$). Drivers who read the OM ($U = 43.00$, $z = -3.042$, $p = .002$) or underwent the L4DTP ($U = 33.00$, $z = -3.520$, $p < .001$) made more correct decisions about whether it was safe and appropriate to activate the automation on the simulator task, compared to drivers who received NT. However, there was no significant difference in the total number of correct decisions made between those who read the OM and those who underwent the L4DTP ($U = 109.50$, $z =$

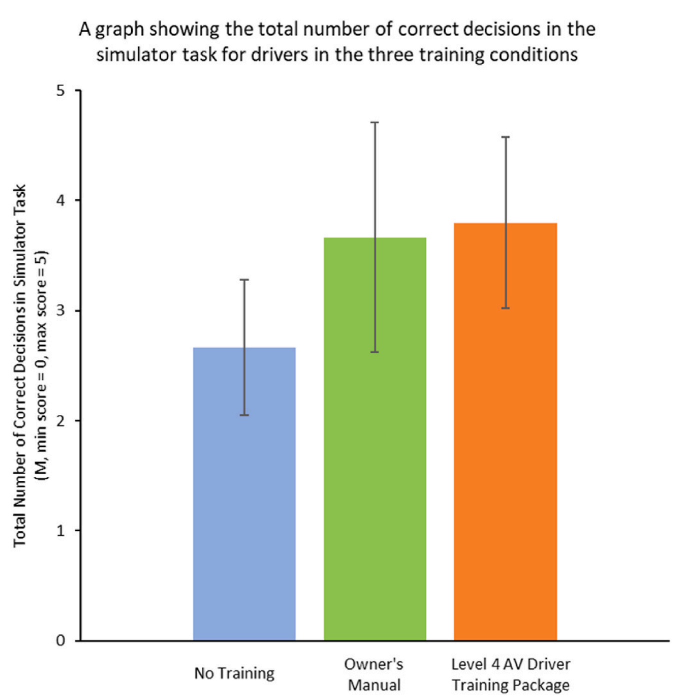


Fig. 14. The effect of type of training on the total number of correct decisions made in the simulator task (error bars represent standard deviations).

−0.133, p = .894; Fig. 14).

3.2. Signal Detection Theory

There was no significant difference in the number of hits (H(2) = 0.89, p = .643) and misses (H(2) = 0.89, p = .643) that drivers committed on the simulator task between the three training conditions. However, there was a significant difference in the number of false alarms (H(2) = 15.91, p < .001) and correct rejections (H(2) = 15.91, p < .001) that drivers committed on the simulator task between the three training conditions. Drivers who read the OM committed fewer false alarms (U = 35.00, z = − 3.360, p = .001) and more correct rejections (U = 35.00, z = − 3.360, p = .001) on the simulator task compared to drivers who received NT. Similarly, drivers who underwent the L4DTP committed fewer false alarms (U = 33.00, z = − 3.487, p < .001) and more correct rejections (U = 33.00, z = − 3.487, p < .001) on the simulator task compared to drivers who received NT. However, there was no significant difference in the number of false alarms (U = 106.50, z = − 0.271, p = .787) and correct rejections (U = 106.50, z = − 0.271, p = .787) that drivers committed on the simulator task between those who read the OM and those who underwent the L4DTP (Fig. 15).

3.3. Trust in Automation

Generally, drivers who received NT showed an increased trust in automation after completing the simulator task (M = 3.60, SD = 6.37). In contrast, drivers who read the OM (M = − 1.50, SD = 12.61) or underwent the L4DTP (M = − 0.33, SD = 10.29) showed reduced trust in automation after undergoing training and completing the simulator task. However, there was no significant difference in the change in drivers' trust in automation between the three training conditions (H(2) = 1.11, p = .575).

3.4. Workload

There was no significant difference in the total workload experienced by drivers on the simulator task between the three training conditions (F (2, 42) = 1.82, p = .174). Similarly, there was no significant difference in the physical demand (H(2) = 0.84, p = .657), temporal demand (F(2, 42) = 0.76, p = .475), performance (F(2, 42) = 1.11, p = .340), effort (H (2) = 3.22, p = .200) or frustration (H(2) = 0.62, p = .735) experienced by drivers on the simulator task between the three training conditions.

However, there was a significant difference in the mental demand experienced by drivers on the simulator task between the three training conditions (H(2) = 7.58, p = .023). Drivers who read the OM experienced a greater mental demand when performing the simulator task compared to drivers who underwent the L4DTP (U = 48.00, z = − 2.685, p = .007; Fig. 16). However, there was no significant difference in the mental demand experienced by drivers on simulator task between drivers who received NT and drivers who read the OM (U = 73.50, z = − 1.625, p = .104) or between drivers who received NT and drivers who underwent the L4DTP (U = 84.00, z = − 1.19, p = .235).

3.5. Mental Models

3.5.1. Appropriate 1

Two drivers in the NT condition, four drivers in the OM condition and two drivers in the L4DTP condition incorrectly decided that it was unsafe to activate the automation in this scenario (Fig. 17). The NT and OM drivers cited the behaviour of other vehicles (e.g. "under-cutting cars", "learner driver", "erratic/unpredictable vehicles") and heavy traffic levels in their reasoning for their decision. However, these variables are not classed as unsafe road conditions (see Fig. 1). In contrast, the L4DTP drivers cited road conditions which were unsafe for activation ("unclear road markings", "thunder", Fig. 18a). These two road conditions were not actually present in the scenario, so although the two drivers in the L4DTP condition made the wrong decision, as these road conditions were unsafe, their reasoning (mental model) for why they made the wrong decision was correct.

Thirteen drivers in the NT condition, 11 drivers in the OM condition and 13 drivers in the L4DTP condition correctly decided that it was safe to activate the automation in this scenario. All drivers mentioned appropriate/safe road conditions in their reasoning for their decision (e.g. "motorway", "speed between 50 and 70 mph", "good weather", "clear lane markings", "good road surface", "straight road"), however the L4DTP group mentioned more appropriate road conditions compared to the OM group and the OM group mentioned more appropriate road conditions compared to the NT group (Fig. 18a). For example, four drivers in the L4DTP group and four drivers in the OM group mentioned

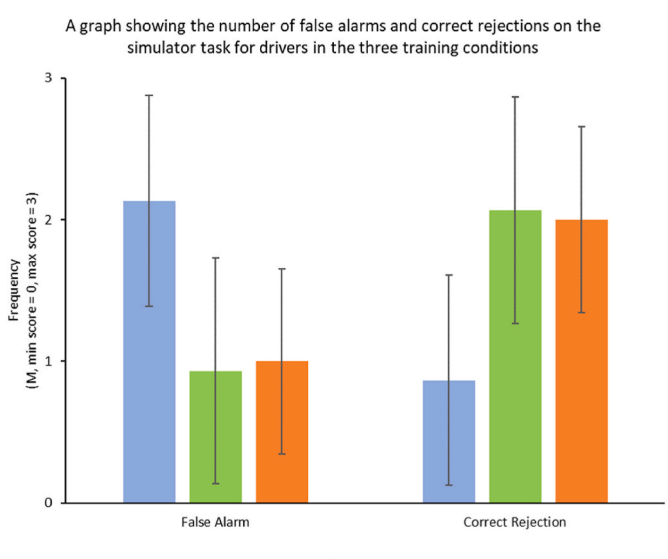


Fig. 15. The effect of type of training on the number of false alarms and correct rejections that were committed on the simulator task (error bars represent standard deviations).

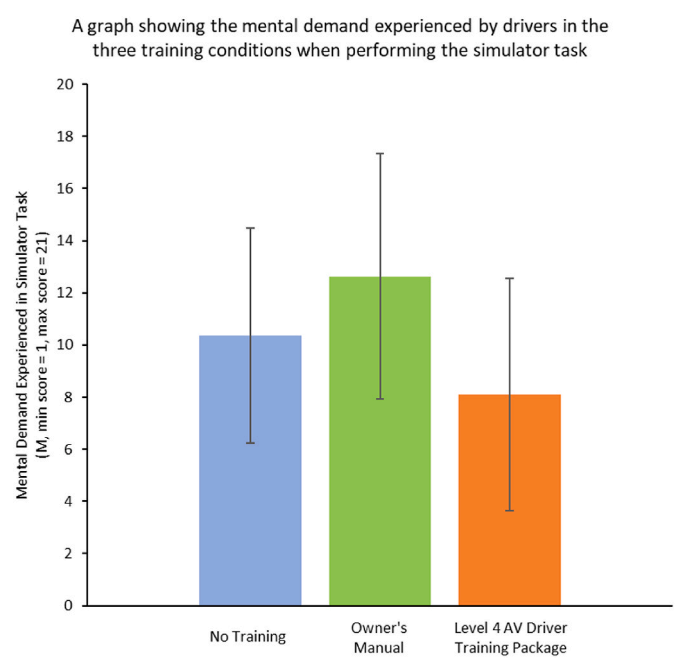


Fig. 16. The mental demand experienced by drivers when performing the simulator task in the three training conditions (error bars represent standard deviations).

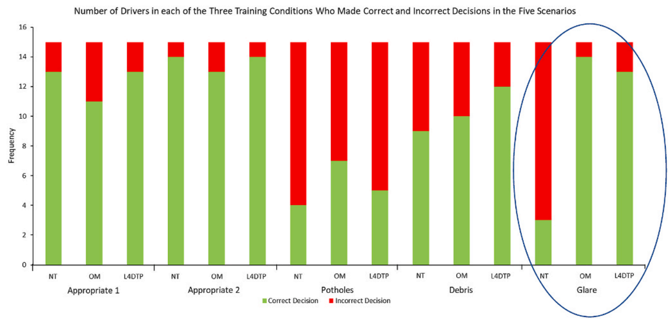


Fig. 17. The number of drivers in each training condition who made correct (green) and incorrect (red) decisions in the five scenarios. The biggest effect could be seen in the glare scenario (circled). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

that the road type was appropriate for activation ("motorway", "non-residential"), however no drivers mentioned this environmental category in the NT group. Similarly, only two drivers in the NT group mentioned appropriate weather conditions and road geometry (e.g. "no rain", "straight"). In comparison, six and four drivers respectively in the OM condition and ten and six drivers respectively in the L4DTP condition cited these environmental categories in their answers (e.g. "good weather/visibility/lighting", "no rain", "clear", "straight road", "slight bend", "no twists/bends"). Additionally, two drivers in the NT group did not offer any reasoning for why it was safe to activate the automation (they answered "it was safe").

3.5.2. Appropriate 2

One driver in the NT condition, two drivers in the OM condition and one driver in the L4DTP condition incorrectly decided that it was unsafe to activate the automation in this scenario (Fig. 17). All drivers cited heavy traffic levels in their reasoning for their decision, even though this is not an inappropriate/unsafe road condition (see Fig. 1). However, the driver in the L4DTP condition also acknowledged that the speed limit (70 mph) was appropriate for activation.

Fourteen drivers in the NT condition, 13 drivers in the OM condition and 14 drivers in the L4DTP condition correctly decided that it was safe to activate the automation in this scenario. All drivers cited appropriate road conditions in their reasoning for their decision (e.g. "motorway", "good speed", "steady/moving traffic", "no obstacles", "good visibility/weather", "no sharp bends", "clear/simple lane markings", "straight road"), however the L4DTP group mentioned more appropriate road conditions compared to the OM group and the OM group mentioned more appropriate road conditions compared to the NT group (Fig. 18b). For example, with the weather conditions category, only two drivers in the NT condition mentioned that the weather conditions were appropriate for activation ("no rain", "good weather"). This rose to four drivers in the OM condition ("good visuals", "no rain or bad weather/sun glare", "good weather") and 13 drivers in the L4DTP condition (e.g. "good visibility/weather", "no rain", "clear day"). Additionally, the OM group tended to focus on the speed category when making their decision (e.g. 10 of the 13 drivers referred to this category compared to three for road geometry and four for road type). The NT group tended to focus on

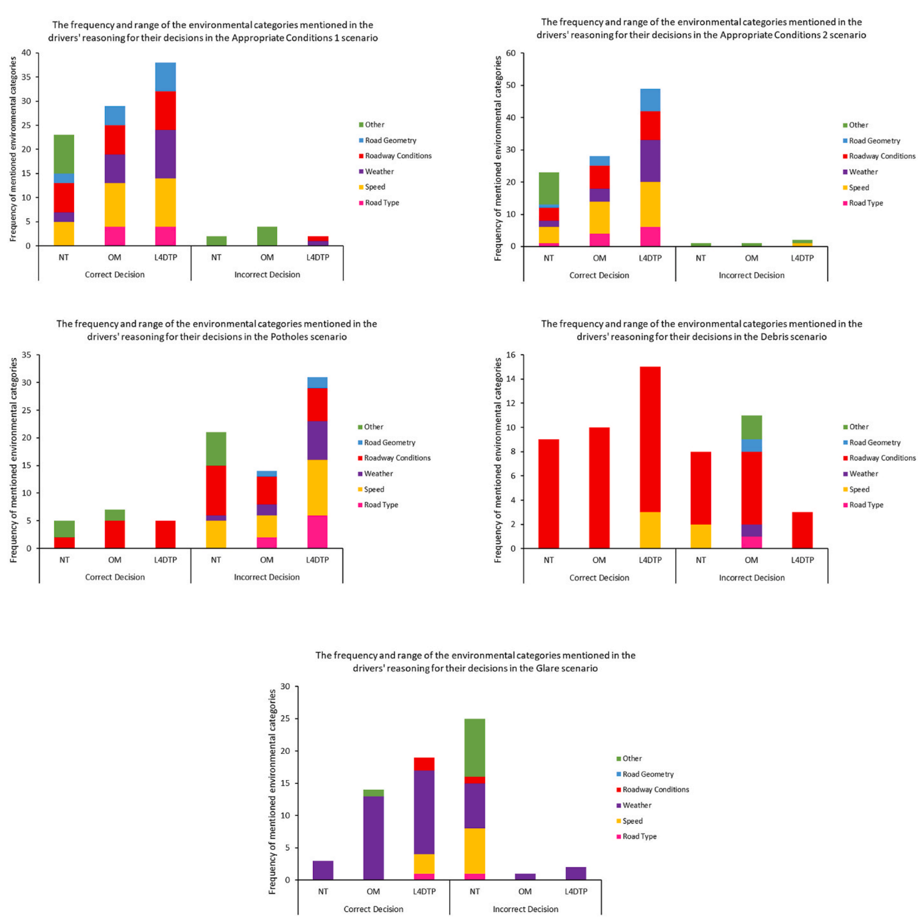


Fig. 18. The frequency and range of the environmental categories that were mentioned in the drivers' reasoning for their correct or incorrect decisions in the Appropriate Conditions 1 (a), Appropriate Conditions 2 (b), Potholes (c), Debris (d) and Glare (e) scenarios. In the two appropriate scenarios, the L4DTP group showed a greater frequency and spread of appropriate road conditions in their reasoning, compared to drivers who received NT or read the OM.

the traffic levels and behaviour of other vehicles (e.g. “busy but moving well”, “no overtaking cars”, “good distance between cars”, n = 10). In comparison, the L4DTP group mentioned a broader range of appropriate road conditions across the five environmental categories (e.g. 13 drivers referred to the weather and speed categories and nine drivers referred to roadway conditions).

3.5.3. Potholes

Eleven drivers in the NT condition, eight drivers in the OM condition and 10 drivers in the L4DTP condition incorrectly decided that it was safe to activate the automation in this scenario (Fig. 17). The 10 drivers in the L4DTP condition did not mention the poor road surface in their reasoning for their decision. In comparison, one driver in the OM condition and seven drivers in the NT condition mentioned the poor road surface (e.g. “small cracks”, “unevenness”, “skid marks”, “damaged road surface”), but decided that it was safe to activate the automation because they thought that the automation would perform better than them (n = 1) or that the poor road surface would not affect the automation. As a poor road surface is an inappropriate road condition (see section 2.2.5.2, Fig. 1), these drivers did not have an appropriate mental model for when the automation can be activated.

The remaining drivers who made the wrong decision (n = 21) did not mention the poor road surface in their reasoning for their decision. This suggests that they either did not see the potholes/cracks on the road or they saw them but did not think it was an issue for the automation. Instead, these drivers mentioned road conditions which were safe for activation (e.g. “motorway”, “speed between 50 and 70 mph”, “smooth”, “dry road”, “good/appropriate weather”, “clear/good markings”, “straight road”, Fig. 18c). If the drivers did not see the potholes/cracks, their reasoning suggests that they did have an appropriate mental model for when the automation can be activated, as they made the wrong decision but used the correct reasoning. If however, they saw the cracks/potholes but did not think it was an issue for the automation, their reasoning suggests that they did not have an appropriate mental model. Based on the evidence presented here, it cannot be said which explanation is correct, demonstrating the need for eye-tracking to be used in future experiments to provide more clarity on whether drivers noticed the poor road surface or not and subsequently whether this is suggestive of an appropriate or inappropriate mental model.

Four drivers in the NT condition, seven drivers in the OM condition and five drivers in the L4DTP condition correctly decided that it was unsafe to activate the automation in this scenario. All drivers in the L4DTP condition correctly identified the poor road surface as the reason for why it was unsafe to activate the automation (e.g. “potholes”, “road bumps”, “cracked road surface”). In comparison, only five drivers in the OM condition and two drivers in the NT condition correctly identified the poor road surface as the reason for their decision (e.g. “oil spillage”, “uneven/poor/cracked road surface”). The remaining four drivers did not mention the poor road surface in their reasoning; they focussed on the traffic levels and the behaviour of other vehicles (e.g. “wanting to stay in fast lane”, “too many cars”, “learner driver on road”). As such, these four drivers did not have an appropriate mental model for when the automation can be activated.

3.5.4. Debris

Six drivers in the NT condition, five drivers in the OM condition and three drivers in the L4DTP condition incorrectly decided that it was safe to activate the automation in this scenario (Fig. 17). All drivers (apart from one in the OM condition) identified the first debris/obstacle as an unsafe road condition (e.g. “obstruction”, “hazard”, “tyre”, “debris”), however all but two of the drivers decided that it was safe to activate the automation once they had passed the obstacle (e.g. “once passed obstruction”). Therefore, these drivers made the wrong decision because they did not anticipate the presence of future obstacles nor wait to see if there were further obstacles in the road ahead. The remaining two drivers (both in the NT condition) decided that it was safe to activate the automation because they thought the automation would be able to deal with the hazard just like a human driver (e.g. “if I can do it so can the automation”). As debris/obstructions are inappropriate road conditions (see section 2.2.5.2, Fig. 1), these two NT drivers did not have an appropriate mental model for when the automation can be activated.

Nine drivers in the NT condition, 10 drivers in the OM condition and 12 drivers in the L4DTP condition correctly decided that it was unsafe to activate the automation in this scenario. All drivers correctly identified the debris/obstacles in their reasoning for their decision (e.g. “multiple/numerous hazards”, “obstacles”, “debris”, “road hazards”, “tyre”, “road work debris”, Fig. 18d). Unlike the drivers who made the wrong decision, these drivers anticipated the potential for future hazards (e.g. “possible future hazards”, “more debris could be around”, n = 2) or mentioned more than one hazard in their reasoning (e.g. “multiple/numerous/too many hazards”, “hazards/obstacles/debris”), suggesting that they waited to see whether there were more obstacles on the road before making their decision. Additionally, some drivers explained why the debris/obstacles were unsafe road conditions (e.g. “slow down fast”, “change lanes regularly”, “required human input”, “required focus/alertness/attention”, “may not be detected by sensors”, “life or death situation”).

3.5.5. Glare

Twelve drivers in the NT condition, one driver in the OM condition and two drivers in the L4DTP condition incorrectly decided that it was safe to activate the automation in this scenario (Fig. 17). All but five of the drivers identified the poor weather conditions in their reasoning for their decision (e.g. “glare”, “reduced visibility”, “fog”, “bright conditions”, Fig. 18e). However, the two L4DTP drivers and four of the seven NT drivers incorrectly decided that it was safe to activate the automation, because they thought that the glare was not too bright and their visibility was ok, so the automation should be fine (e.g. “if I can see, so can the automation”). In comparison, the OM driver and three drivers in the NT condition incorrectly decided to activate the automation, because they believed that the glare would not affect the automation (n = 2) or the automation would perform better than them (e.g. “automation sensors would be able to see better”, “automation would be safer than my impaired view”, “would trust automation over myself”). As glare and fog are inappropriate road conditions (see section 2.2.5.2, Fig. 1), these drivers did not have an appropriate mental model for when the automation can be activated.

The remaining five drivers (all in the NT group) did not cite the poor weather conditions in their reasoning for their decision, suggesting that they did not see the glare or think that it was a hazard for the AV system. Although some of these drivers referred to appropriate road conditions in their reasoning for why it was safe to activate the automation (e.g. “motorway”, “constant speed”, “no obstacles”, see Fig. 18e), most drivers based their decision on the traffic levels and the behaviour of other vehicles (e.g. “not too busy”, “no changing lanes”, “large gap/space between vehicles”).

Three drivers in the NT condition, 14 drivers in the OM condition and 13 drivers in the L4DTP condition correctly decided that it was unsafe to activate the automation in this scenario. All drivers (apart from one in the OM condition) correctly identified the poor weather conditions in their reasoning for why it was unsafe to activate the automation (e.g. “fog”, “glare”, “poor lighting”, “reduced/poor visibility”, “low sunshine”, “bright light”). The driver in the OM condition decided that it was unsafe to activate the automation because it was “too busy” on the road.

4. Discussion

AV systems are expected to improve road safety by reducing the frequency and severity of on-road collisions (Schoettle and Sivak, 2014). However, if drivers have an inappropriate mental model for the capabilities and limitations of the automation, they may over-trust and activate the automation in inappropriate road conditions (Lee and See, 2004; Korber et al., 2018), which could cause a collision (Merriman et al., 2021b). As such, this article describes the development of a training package for the safe activation of an AV system (L4DTP) and evaluates its effectiveness in a between-subjects simulator experiment. Drivers received NT (control group), read an OM (experimental group 1: current training provision) or underwent the L4DTP (experimental group 2: new training programme). All drivers then experienced five scenarios in a driving simulator where they encountered road conditions which were safe and unsafe for activation. Their activation decisions, behaviour, trust in automation, workload and mental models were measured.

This evaluation found that drivers who read the OM or underwent the L4DTP made more correct decisions about whether it was safe and appropriate to activate the automation and committed fewer false alarms and more correct rejections on the simulator task compared to drivers who received NT. However there were no differences between drivers who read the OM and drivers who underwent the L4DTP or for the number of hits and misses on the simulator task. These results demonstrate the benefits of undergoing any training for AV systems; if drivers do not receive training, they are more likely to activate the automation in unsafe road conditions (false alarms), which could cause a collision (e.g. Merriman et al., 2021b). For example in the glare scenario, only three drivers in the NT condition correctly decided that it was unsafe to activate the automation, therefore 12 drivers activated the automation in unsafe road conditions. In comparison, only one driver in the OM condition and two drivers in the L4DTP condition activated the automation in the glare scenario (see Fig. 17). Therefore, the OM and L4DTP helped drivers make more appropriate decisions (that it was unsafe to activate the automation) in the unsafe scenarios. In reality, most drivers do not read their OM in a dedicated 45-minute time period and if they do, they do not read all of it (Mehlenbacher et al., 2002). For example, in this evaluation, 68% of drivers reported reading their OM. However the majority of these drivers (57%) had read less than half and only 10% had read the whole manual. Therefore, it could be argued that the comparison between NT and the L4DTP (rather than between the OM and the L4DTP) is the more realistic comparison, and this comparison shows why additional mandatory training for AV systems is needed. If drivers do not read their full OM (i.e. receive NT), they are more likely to activate the automation in unsafe road conditions, which could cause a collision (Merriman et al., 2021b). Therefore these results support previous training studies in showing that some training for AV systems is better than having no training at all (e.g. Ebnali et al., 2019) and extends these studies by showing that AV driver training programmes can also improve drivers' activation decisions and behaviour (previous studies have mostly evaluated drivers' takeover decisions and behaviour e.g. Boelhouwer et al., 2019; Sportillo et al., 2019; Krampell et al., 2020).

Additionally, even if drivers read the OM, the workload and mental model results suggest that it is more beneficial for drivers to undergo the L4DTP. This evaluation found that drivers who underwent the L4DTP experienced a lower mental demand when performing the simulator task compared to drivers who read the OM. Therefore, drivers in the L4DTP condition found it easier, simpler and less demanding (Hart and Staveland, 1988) to decide whether it was safe and appropriate to activate the automation (and act accordingly) compared to drivers in the OM condition. Although not significant, this evaluation also found that drivers in the L4DTP condition experienced a lower total workload, physical demand, temporal demand and effort when performing the simulator task compared to drivers in the other two training conditions. Therefore, drivers in the L4DTP condition found the simulator task less demanding and strenuous and felt under less time pressure when making their decisions compared to drivers who received NT and drivers who read the OM (Hart and Staveland, 1988) and they had to work less hard and put less effort into the simulator task to achieve the same performance as those who read the OM. This suggests that it was easier to learn, remember and apply the knowledge learnt in the L4DTP compared to the OM. These findings support previous training studies in showing that an OM may not be the most effective training method for AV systems and alternative methods such as behavioural training (e.g. Shaw et al., 2020), video-based training (Cahour and Forzy, 2009), a combination of OM and simulator training (e.g. Hergeth et al., 2017; Koustanaï et al., 2012; Krampell et al., 2020) or a combination of written descriptions and video-based training (this study) are needed.

Finally, a qualitative analysis was performed on the drivers' reasoning for their decisions. In the safe scenarios (Appropriate 1 and Appropriate 2), this analysis revealed that most drivers in the L4DTP condition made the wrong decisions because they thought they noticed unsafe road conditions in the scenarios (thunder, unclear lane markings). Therefore, although their decisions were wrong, their reasoning for why they made the wrong decision was correct. This demonstrates that they were incorporating the training into their decision-making processes. In comparison, the OM and NT drivers focussed on the traffic levels and the behaviour of other vehicles, both of which are not unsafe road conditions (see Fig. 1). This suggests that the L4DTP drivers had more appropriate mental models for when the automation can be activated compared to the drivers in the other two training conditions. With regards to the correct decisions, the analysis revealed that more drivers in the L4DTP condition cited appropriate road conditions in their reasoning for their decisions and they mentioned more appropriate road conditions across the five environmental categories compared to drivers in the other two training conditions. In comparison, the OM drivers tended to focus on the speed category in their reasoning for their decisions and the NT drivers tended to focus on the traffic levels and the behaviour of other vehicles. Therefore in the safe scenarios, these findings suggest that drivers in the L4DTP condition had more appropriate, complete and comprehensive mental models for when the automation can be activated compared to drivers in the other two training conditions.

In the unsafe scenarios (Potholes, Debris, Glare), the analysis revealed that most drivers in the OM and L4DTP conditions made the wrong decisions because they did not see the hazard or anticipate the presence of future hazards. In comparison, most drivers in the NT condition made the wrong decisions because they thought that the hazards would not affect the automation, or the automation would perform better than them. With regards to making the correct decisions, this analysis found that all drivers in the L4DTP condition and most drivers in the OM and NT conditions made the correct decisions because they identified the hazard and used this as reasoning for why it was unsafe to activate the automation. Therefore in the unsafe scenarios, drivers who read the OM or underwent the L4DTP had more appropriate mental models for when the automation can be activated compared to drivers who received NT.

4.1. Evaluation and future work

There are limitations with this training evaluation. Firstly, the selection of the simulator scenarios was based upon the video-clip scenarios which drivers found most difficult in the previous online evaluation (Merriman et al., Under Review). Although five scenarios were identified, simulator issues meant that the poor lane markings and sharp bend scenarios could not be created or tested in this evaluation. As numerous drivers made errors on these scenarios in the previous online evaluation, future research should investigate whether the L4DTP improves drivers' activation decisions and behaviour in these scenarios. Additionally, all drivers were given 45 minutes to complete their allocated training programme (OM or L4DTP). However, the training programmes involved different methods of delivery (reading vs interacting with the L4DTP). Most OM drivers did not spend the full 45 minutes reading the OM, therefore there may have been differences between groups in their cognitive fatigue which in-turn may have affected their performance on the simulator task. This could be investigated in future research.

Secondly, the L4DTP combined an acronym with an online video-based training module. Although this training evaluation demonstrated benefits of the training package over NT and an OM, it is unclear whether the acronym, the online training module or both components were important and responsible for improving drivers' decisions, mental demand, mental models and behaviour. Therefore, future work should investigate each component in isolation and together to investigate which combination is most effective in improving drivers' behaviour. Additionally, this training evaluation did not measure the differences between groups in their pre- and post-training scores (apart from the variable Trust in Automation), instead looking only at the differences between groups in their post-training scores. This means that all groups could not be shown to be equivalent in their decisions, workload, mental models and behaviour before training, which makes it hard to conclude that any group differences that were found after training was due to the training that drivers received (it could be due to pre-existing group differences). The difference between pre- and post-training scores were not measured for all variables because of the time and resource constraints that were present due to the study taking place at a time when in-person participant testing was harder with covid restrictions. As such, a trade-off had to be made. Trust in Automation was considered to be an important variable in this study, hence why the difference between pre- and post-training trust scores was calculated. However, future research should measure pre- and post-training scores for all variables to investigate whether there are any group differences in drivers' decisions, workload, mental models and behaviour before training. This means that if no differences are found, researchers can more confidently conclude that any group differences that are found after training are due to the training that drivers received.

Thirdly, the evaluation took place in a driving simulator. Although simulators have the advantage of allowing researchers to standardise, control and reproduce the same scenarios for each driver (e.g. traffic, weather, road layout), the consequences of making mistakes are less severe than in real life as drivers are not placed into situations where they may suffer physical harm (De Winter et al., 2012; Harvey and Burnett, 2019). In this evaluation, if drivers incorrectly decided that it was safe to activate the automation in the unsafe scenarios, the only consequence was the lack of marks that they received. However, if drivers make the same poor decisions on real life roads, they could cause a collision resulting in fatal consequences (e.g. Merriman et al., 2021b). This lower perception of risk may have caused drivers to become riskier and activate the automation in more scenarios than they would actually activate the automation in real life. This limits the generalisability of the findings to the real world as drivers' simulated driving behaviour may not reflect their actual driving behaviour. Therefore, future on-road evaluations are needed to investigate whether the L4DTP improves drivers' activation decisions and behaviour on the road.

The effectiveness of the L4DTP was only evaluated in the short-term (immediately after training). The workload results suggest that drivers in the OM condition found it harder to remember and apply their newly learnt knowledge when making their decisions. Over time these drivers may forget the information in the OM, therefore there could be differences in the appropriateness and comprehensiveness of drivers' mental models, their activation decisions and behaviour if evaluated in the long-term (e.g. after a few months, a year). The OM is a passive learning method as drivers just read the information in the manual (Bell and Kozlowski, 2008). In contrast, the L4DTP blends passive and active training methods as drivers are actively involved in the learning process by reading information, watching video-clips, answering questions and receiving feedback. A recent study suggests that active learning methods may have a greater retention rate compared to passive learning methods after a month's delay (Minnick et al., 2022) and as drivers are unlikely to read their OM more than once, this study suggests that even if drivers remember the content in the OM immediately after training, it may degrade one month later. Therefore, there could be long-term retention benefits of the L4DTP over the OM which should be investigated in future longitudinal evaluations.

Finally, for the purpose of assigning drivers' decisions into "correct" or "incorrect" categories, road and environmental conditions were categorised as "safe" or "unsafe", based upon the ODD defined in Fig. 1. These environmental conditions then formed the content of the training programmes in order to gain insights into training comprehension and retention. However, the mental model results show that there were some influences on drivers' decisions, that were not part of the ODD, which caused drivers to make a different decision than what was originally planned. For example, heavy traffic and the erratic behaviour of other drivers were not defined as "unsafe" road conditions, so they were not included as unsafe road conditions in the training programmes. However, quite a few drivers did not activate the automation in the safe/appropriate scenarios for these reasons (see sections 3.5.1 and 3.5.2). Due to the categorisation of the "safe" and "unsafe" road conditions, these decisions were marked as incorrect and drivers' reasoning (mental models) was deemed inappropriate. However, these are unsafe road conditions, and it is safer for drivers to act more cautiously on the road and not activate the automation whenever they consider the road conditions to be unsafe, regardless of whether those road conditions are actually defined as unsafe according to the AV system's ODD. This suggests that in future AV driver training programmes, drivers should not just be taught which road conditions are safe and unsafe, but also to be cautious and to not activate the automation whenever they consider the road environment to be unsafe. Similarly in training evaluations, rather than marking drivers down and assigning decisions like these as incorrect decisions, there should be a middle category where the decisions are correct because the driver considered the road conditions to be unsafe, even though the road conditions were not actually defined as unsafe according to the AV system's ODD. This also highlights the need for future AV system design to always allow drivers to make the final decision on whether to activate the automation.

This study is one of the first studies to develop a training programme for the safe activation of an AV system and evaluate its effectiveness through quantitative and qualitative methodologies. In the AV driver training literature, the bulk of research has used quantitative methods to evaluate the effectiveness of driver training programmes (e.g. number of collisions, takeover time, standard deviation of lateral position, accelerator pressure, brake pressure, takeover behaviours: see Merriman et al., 2021a). However in this evaluation, quantitative and qualitative analyses were performed. Although the quantitative results revealed some benefits of the L4DTP over NT (correct decisions, correct rejections, false alarms) and OMs (workload), the qualitative results revealed more reasons for why the L4DTP was more effective than the two other training conditions (more appropriate and comprehensive mental models). Therefore, this evaluation adds further evidence for why a mixed methods approach to data collection should be used.

Although this article focussed on one automation task (activation), one AV system and made various assumptions about the road conditions which the AV system can safely operate in (e.g. glare, see section 2.2.5.2), this evaluation showed that a blended learning methodology, combining video-based online training and an acronym, is effective and can be used to improve drivers' mental models, activation decisions, behaviour and workload, regardless of the tasks, systems and assumptions that are made. Therefore when other levels of automation, other AV systems or other tasks (e.g. takeovers) are investigated or the technology improves and the assumptions need to change (e.g. the automation can now cope with glare but cannot deal with a particular road sign), the same methodology can be used to train drivers; it is only the training content that will need to be changed (e.g. video-clips of those road signs could be added to the online video-based component of the L4DTP).

5. Conclusion

AV systems are expected to improve road safety. Unless drivers have an appropriate mental model for the capabilities and limitations of the automation, they may over-trust and activate the automation in inappropriate road conditions, potentially leading to a collision. Therefore, training in the appropriate activation of AV systems is essential. A L4DTP was developed to improve drivers' decisions and behaviour when activating an AV system. A between-subjects simulator experiment showed that drivers who read the OM or underwent the L4DTP made better activation decisions and showed better activation behaviour (fewer false alarms and more correct rejections) compared to drivers who received NT. Although all drivers are given an OM to read, in reality most drivers do not read their full OM, therefore the NT results reflect what may occur if drivers are only given an OM to read and demonstrates the need for additional mandatory training for AV systems. Additionally, drivers who underwent the L4DTP found it easier, less demanding and felt under less time pressure when making their decisions, had to expend less effort to reach the same activation performance and had more appropriate and comprehensive mental models for when the automation can be activated compared to drivers who read the OM. This L4DTP can make roads safer by reducing collisions linked to poor activation decisions and behaviour. Therefore if this training programme is adopted into mandatory AV driver training, the safety benefits of AV systems can be realised.

Funding

This research was funded by IAM RoadSmart and the Engineering and Physical Sciences Research Council. These funders had no involvement in the study design, in the collection, analysis, and interpretation of the data, in the writing of the report, and in the decision to submit this paper for publication.

Declaration of competing interest

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

References

Audi, 2017. Audi Elaine Concept Car – Highly Automated at Level 4. November 09. Retrieved May 20, 2020, from Audi MediaCenter: https://www.audi-mediacenter.com/en/on-autopilot-into-the-future-the-audi-vision-of-autonomous-driving-9305/audi-elaine-concept-car-highly-automated-at-level-4-9308.

Bagloee, S.A., Tavana, M., Asadi, M., Oliver, T., 2016. Autonomous vehicles: challenges, opportunities, and future implications for transportation policies. J. Mod. Transport. 24, 284–303.

Barg-Walkow, L.H., Rogers, W.A., 2016. The effect of incorrect reliability information on expectations, perceptions, and use of automation. Hum. Factors 58 (2), 242–260.

Beggiato, M., Krems, J.F., 2013. The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information. Transport. Res. F Traffic Psychol. Behav. 18, 47–57.

Beggiato, M., Pereira, M., Petzoldt, T., Krems, J.F., 2015. Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study. Transport. Res. F Traffic Psychol. Behav. 35, 75–84.

Bell, B.S., Kozlowski, S.W., 2008. Active learning: effects of core training design elements on self-regulatory processes, learning, and adaptability. J. Appl. Psychol. 93 (2), 296–316.

Bellezza, F.S., 1981. Mnemonic devices: classification, characteristics, and criteria. Rev. Educ. Res. 51, 247–275.

Blömacher, K., Nöcker, G., Huff, M., 2020. The evolution of mental models in relation to initial information while driving automated. Transport. Res. F Traffic Psychol. Behav. 68, 198–217.

Boelhouwer, A., van den Beukel, A.P., van der Voort, M.C., Martens, M.H., 2019. Should I take over? Does system knowledge help drivers in making take-over decisions while driving a partially automated car? Transport. Res. F Traffic Psychol. Behav. 60, 669–684.

Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qual. Res. Psychol. 3 (2), 77–101.

Cahour, B., Forzy, J.-F., 2009. Does projection into use improve trust and exploration? An example with a cruise control system. Saf. Sci. 47 (9), 1260–1270.

Casner, S.M., Hutchins, E.L., 2019. What do we tell the drivers? Toward minimum driver training standards for partially automated cars. J. Cogn. Eng. Decis. Mak. 13 (2), 55–66.

Choi, J.K., Ji, Y.G., 2015. Investigating the importance of trust on adopting an autonomous vehicle. Int. J. Hum. Comput. Interact. 31 (10), 692–702.

Cruise, C., 2018. Human Error Prediction in Automated Driving [Unpublished Master's Thesis]. University of Southampton, Southampton, UK.

Daimler, 2019. Daimler and BMW Group: Long-Term Development Cooperation for Automated Driving. July 04. Retrieved May 21, 2020, from Product Innovation: https://www.daimler.com/innovation/case/autonomous/development-cooperation-daimler-bmw.html.

De Winter, J.C.F., Van Leeuwen, P.M., Happee, R., 2012. Advantages and Disadvantages of Driving Simulators: A Discussion. Proceedings of Measuring Behavior, Utrecht, Netherlands, pp. 47–50.

Degani, A., Wiener, E.L., 1997. Procedures in complex systems: the airline cockpit. IEEE Trans. Syst. Man Cybern. Syst. Hum. 27 (3), 302–312.

Driver and Vehicle Standards Agency, 2014. Car And Light Van Driving Syllabus (Category B). November 24. Retrieved September 07, 2020, from Learning to drive a car syllabus: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/377664/car-and-light-van-driving-syllabus.pdf.

Ebnali, M., Hulme, K., Ebnali-Heidari, A., Mazloumi, A., 2019. How does training effect users' attitudes and skills needed for highly automated driving? Transport. Res. F Traffic Psychol. Behav. 66, 184–195.

Ehrman, M.E., Leaver, B.L., Oxford, R.L., 2003. A Brief Overview of Individual Differences in Second Language Learning. System, p. 31.

Evans, R.L., 2007. Every Good Boy Deserves Fudge: the Book of Mnemonic Devices. A Perigee Book, The Penguin Group, NewYork, USA.

Ford Motor Company, 2016. Ford of the Future: Autonomous Vehicles and Silicon Valley Investments | Innovation | Ford. August 16. from. https://www.youtube.com/watch?v=dDiHtlI0OhI. (Accessed 13 May 2020).

Forster, Y., Hergeth, S., Naujoks, F., Krems, J., Keinath, A., 2019. User education in automated driving: owner's manual and interactive tutorial support mental model formation and human-automation interaction. Information 10 (4), 1–22.

Gentner, D., Stevens, A.L., 1983. Mental Models, 1 ed. Psychology Press.

Gibson, H.A., 2009. Using mnemonics to increase knowledge of an organizing curriculum framework. Teach. Learn. Nurs. 4, 56–62.

Gold, C., Körber, M., Hohenberger, C., Lechner, D., Bengler, K., 2015. Trust in automation – before and after the experience of take-over scenarios in a highly automated vehicle. Procedia Manuf. 3, 3025–3032.

Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics, vol. 1. Wiley, New York.

Greenblatt, J.B., Shaheen, S., 2015. Automated vehicles, on-demand mobility, and environmental impacts. Curr. Sustain./Renew. Energy Rep. 2 (3), 74–81.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload. Elsevier Science Publishers, North-Holland, pp. 139–183.

Harvey, C., Burnett, G., 2019. The influence of incentives and instructions on behaviour in driving simulator studies. Transport. Res. F Traffic Psychol. Behav. 67, 164–174.

Hays, R.T., Jacobs, J.W., Prince, C., Salas, E., 1992. Flight simulator training effectiveness: a meta-analysis. Mil. Psychol. 4 (2), 63–74.

Hergeth, S., Lorenz, L., Krems, J.F., 2017. Prior familiarization with takeover requests affects drivers' takeover performance and automation trust. Hum. Factors 59 (3), 457–470.

Higbee, K.L., 2001. Your Memory: How it Works and How to Improve it, second ed. Da Capo Press, Cambridge.

Honda, 2017. Honda Announces New Goal for Highly-Automated Driving Systems [Video File]. June 07. Retrieved May 21, 2020, from. https://www.youtube.com/watch?v=LSAg6bzlf1Q&feature=youtu.be.

Hwang, M.I., 1994. Decision making under time pressure: a model for information systems research. Inf. Manag. 27 (4), 197–203.

Korber, M., Baseler, E., Bengler, K., 2018. Introduction matters: manipulating trust in automation and reliance in automated driving. Appl. Ergon. 66, 18–31.

Koustanaï, A., Cavallo, V., Delhomme, P., Mas, A., 2012. Simulator training with a forward collision warning system: effects on driver-system interactions and driver trust. Hum. Factors 54 (5), 709–721.

Kovar, S.K., Van Pelt, C., 1991. Using a first-letter mnemonic to improve basketball set-shot performance. Percept. Mot. Skills 72, 1383–1390.

Krampell, M., Solís-Marcos, I., Hjälmdahl, M., 2020. Driving automation state-of-mind: using training to instigate rapid mental model development. Appl. Ergon. 83, 102986.

Lee, J.D., See, K.A., 2004. Trust in automation: designing for appropriate reliance. Hum. Factors 46 (1), 50–80.

Lewis, J.B., Mulligan, R., Kraus, N., 2018. The importance of medical mnemonics in medicine. Pharos 30–35.

Manser, M.P., Noble, A.M., Machiani, S.G., Shortz, A., Klauer, S.G., Higgins, L., Ahmadi, A., 2019. Driver Training Research and Guidelines for Automated Vehicle Technology. Safe-D National UTC, Texas A&M Transportation Institute, Virginia Tech Transportation Institute & San Diego State University, Texas, US.

Marks, M.A., Zaccaro, S.J., Mathieu, J.E., 2000. Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. J. Appl. Psychol. 85, 971–986.

McCabe, J.A., Osha, K.L., Roche, J.A., Susser, J.A., 2013. Psychology students' knowledge and use of mnemonics. Teach. Psychol. 40 (3), 183–192.

Mehlenbacher, B., Wogalter, M.S., Laughery, K.R., 2002. On the reading of product owner's manuals: perceptions and product complexity. In: Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, vol. 46, pp. 730–734, 6.

Merriman, S. E., Revell, K. M., & Plant, K. L. (Under Review). What Does an Automated Vehicle Class as a Hazard? Using Online Video-Based Training to Improve Drivers'

Trust and Mental Models for Activating an Automated Vehicle. Manuscript submitted for publication.
Merriman, S.E., Plant, K.L., Revell, K.M., Stanton, N.A., 2021a. Challenges for automated vehicle driver training: a thematic analysis from manual and automated driving. Transport. Res. F Traffic Psychol. Behav. 76, 238–268. https://doi.org/10.1016/j.trf.2020.10.011.
Merriman, S.E., Plant, K.L., Revell, K.M., Stanton, N.A., 2021b. What can we learn from Automated Vehicle collisions? A deductive thematic analysis of five Automated Vehicle collisions. Saf. Sci. 141, 105320 https://doi.org/10.1016/j.ssci.2021.105320.
Merriman, S.E., Plant, K.L., Revell, K.M., Stanton, N.A., 2023. A new approach for training needs analysis: a case study using an automated vehicle. Appl. Ergon. 111, 104014 https://doi.org/10.1016/j.apergo.2023.104014.
Minnick, W., Cekada, T., Marin, L., Zreiqat, M., Seal, B., Mulroy, J., 2022. The impact of active learning strategies on retention and outcomes in safety training. Creativ. Educ. 13 (2), 526–536.
Montag, I., Comrey, A.L., 1987. Internality and externality as correlates of involvement in fatal driving accidents. J. Appl. Psychol. 72 (3), 339–343.
National Transportation Safety Board, 2020. Collision between a Sport Utility Vehicle Operating with Partial Driving Automation and a Crash Attenuator. February 25. Retrieved April 23, 2020, from Accident Report Detail: https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR2001.pdf.
Norman, D.A., 1988. The Psychology of Everyday Things. Basic Books, New York.
Putnam, A.L., 2015. Mnemonics in education: current research and applications. Translat. Issues Psychol. Sci. 1 (2), 130–139.
Radović, T., Manzey, D., 2019. The impact of a mnemonic acronym on learning and performing a procedural task and its resilience toward interruptions. Front. Psychol. 10, 1–17.
IAM RoadSmart, 2016. Advanced Driver Course Associate Logbook. Retrieved December 18, 2019, from IAM RoadSmart: https://pcc-hub.com/iam/p/5491/advanced-driver-course-associate-logbook.
Rowe, A.L., Cooke, N.J., 1995. Measuring mental models: choosing the right tools for the job. Hum. Resour. Dev. Q. 6 (3), 243–255.
Saffarian, M., de Winter, J.C., Happee, R., 2012. Automated driving: human-factors issues and design solutions. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 56 (1), 2296–2300.
Schoettle, B., Sivak, M., 2014. A Survey of Public Opinion about Autonomous and Self-Driving Vehicles in the U.S., the U.K., and Australia. Transportation Research Institute (UMTRI)., Ann Arbor, Michigan. Retrieved from. https://deepblue.lib.umich.edu/bi.
Shaw, E., Large, D.R., Burnett, G., 2020. Driver Training for Future Automated Vehicles: Introducing CHAT (CHeck, Assess and Takeover). RAC Foundation, London from. https://www.racfoundation.org/research/safety/driver-training-for-future-automated-vehicles. (Accessed 16 November 2020).
Skoda, 2018. Five Levels of Autonomous Driving. February 27. Retrieved May 21, 2020, from Skoda Storyboard: https://www.skoda-storyboard.com/en/innovation/five-levels-of-autonomous-driving/.
Society of Automotive Engineers International, 2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. June 15. Retrieved December 05, 2019, from SAE International: https://www.sae.org/standards/content/j3016_201806/.
Sportillo, D., Paljic, A., Ojeda, L., 2019. On-road evaluation of autonomous driving training. In: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, Daegu, South Korea, pp. 1–9.
Stanton, N.A., Young, M.S., 2005. Driver behaviour with adaptive cruise control. Ergonomics 48 (10), 1294–1313.
Stanton, N.A., Revell, K.M., Langdon, P., 2021. Designing Interaction and Interfaces for Automated Vehicles: User-Centred Ecological Design and Testing. CRC Press, Boca Raton, Florida.
Stanton, N.A., Brown, J.W., Revell, K.M., Kim, J., Richardson, J., Langdon, P., Thompson, S., 2022. OESDs in an on-road study of semi-automated vehicle to human driver handovers. Cognit. Technol. Work 24, 317–332.
Tesla, 2019. Model S Owner's Manual. May 16. Retrieved May 17, 2021, from. https://tesla-info.com/doc/2019-tesla-model-s-eu.pdf.
Toyota, 2019. Yaris Owner's Manual OM52A96. May. Retrieved from. https://www.toyota.co.uk/.
Ullius, D., 1997. ART: acronyms reinforce training. Train. Dev. 51 (2), 9.
Volvo Cars, 2017. Volvo Cars CEO Urges Governments and Car Industry to Share Safety-Related Traffic Data. April 03. Retrieved May 13, 2020, from Volvo Cars: https://www.media.volvocars.com/global/en-gb/media/pressreleases/207164/volvo-cars-ceo-urges-governments-and-car-industry-to-share-safety-related-traffic-data.
Zakay, D., Wooler, S., 1984. Time pressure, training, and decision effectiveness. Ergonomics 27 (3), 273–284.