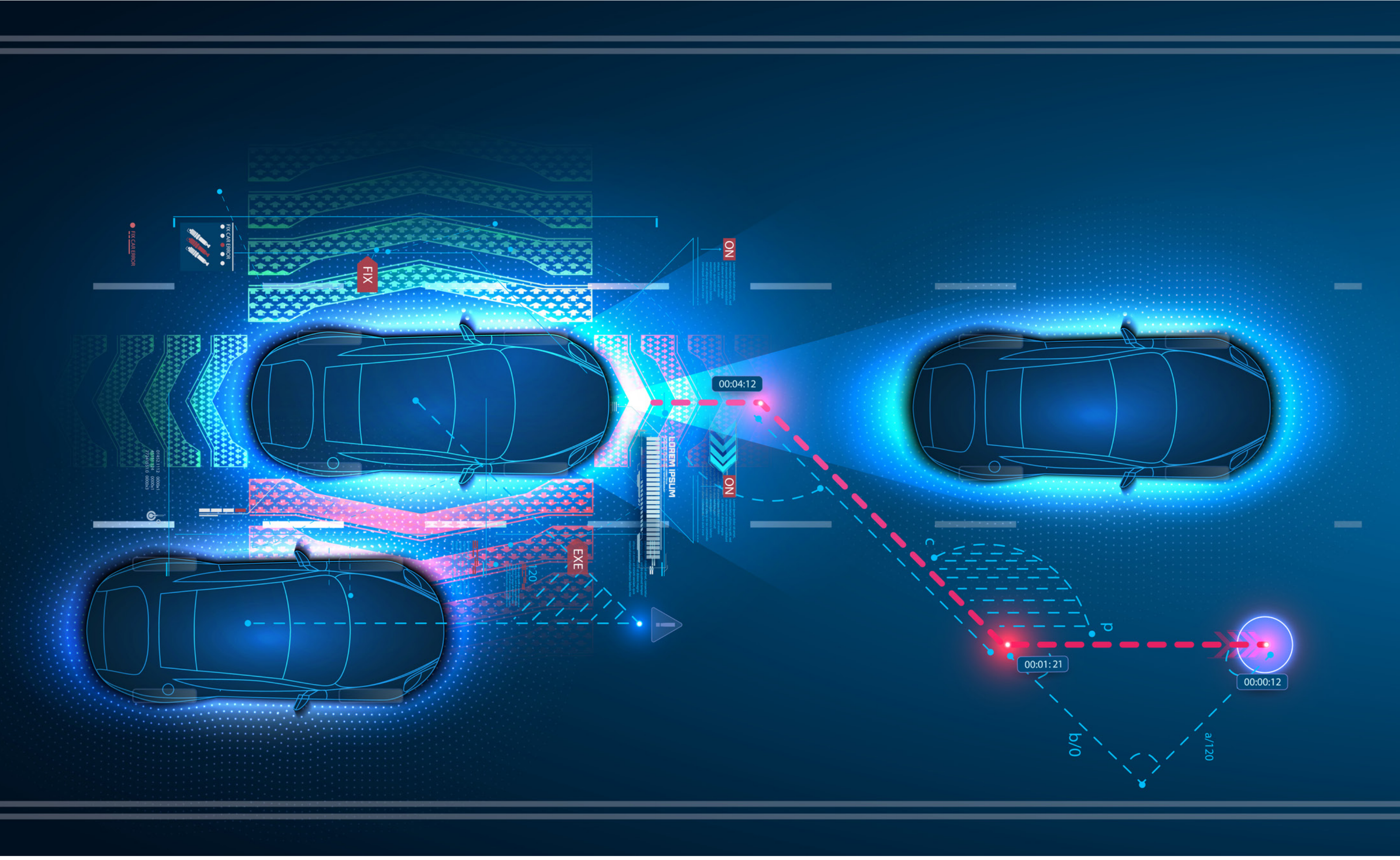


A review of CAV safety benchmarking and a proposal for a "Digital Commentary Driving" technique

June 2021

bsi.

Centre for Connected
& Autonomous Vehicles

Project Team

Lead author: Professor Nick Reed - Reed Mobility
Bryn Balcombe - ADA
Paul Spence - PKSInsights Ltd
Dr. Siddartha Khastgir - WMG, University of Warwick
Nick Fleming - BSI, Head of Sector Transport & Mobility

BSI is grateful to all interview participants for their participation and insights.

About

BSI

BSI helps organizations embed habits of excellence throughout their businesses, enabling them to be resilient and pass the test of time. Our vision is to be the global business improvement partner of choice. BSI is a non-profit distributing company and is proud to hold a Royal Charter. It works in the public interest and is recognized as a leader in standards development. We provide and maintain the infrastructure for experts from industry, government and societal stakeholders to join committees and thus participate in the shaping, development and maintaining of voluntary standards of national, regional and international relevance. The standards that BSI helps shape are developed in response to there being an identified and unmet need for good practice and harmonization. We work with standards bodies around the world to help develop best practice codes and standards, including ISO, IEC, CEN and CENELEC.[1]

About CCAV

The Centre for Connected and Autonomous Vehicles (CCAV) is a joint Department for Business, Energy & Industrial Strategy (BEIS) and Department for Transport (DfT) unit. Established in 2015, CCAV is an expert unit that is working with industry and academia to make everyday journeys greener, safer, more flexible and more reliable by shaping the safe and secure emergence of connected and self-driving vehicles in the UK and by leading the government's Future of Transport strategy.

Important Notice

This report has been prepared for general information purposes relating to its subject matter only. It does not constitute a definitive or final advice; it outlines a number of possible courses of action and next steps and is intended to inform further stakeholder discussion and decisions on its subject.

For more information on its subject matter specifically, please contact CAV@bsigroup.com. More information on the BSI CAV Programme can be found here: www.bsigroup.com/cav.

---

[1] ISO, International Organization for Standardization; IEC, International Electrotechnical Commission; CEN, European Committee for Standardization; CENELEC, European Committee for Electrotechnical Standardization

Contents

List of Figures

List of Tables

Executive summary

This pre-standardization research project explored approaches to assessing the safety performance of connected and automated vehicles (CAVs) that would be considered acceptable both in terms of overall safety and from the perspective of other road users. Such assessments were recognized in the context of the operational design domain (ODD) in which the CAV has been designed to operate, as specified within PAS 1883. The focus was on the behaviours and outcomes of complete CAVs for use in service (rather than for testing), agnostic to the sensors and algorithms used.

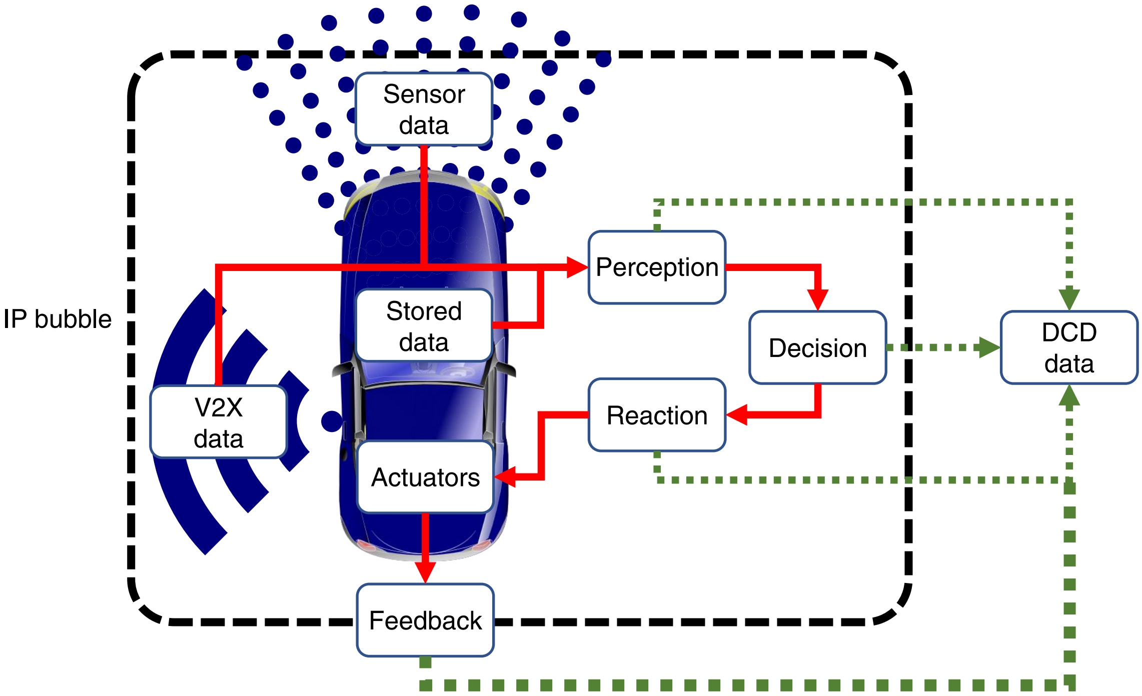
The research was comprised of two phases – a literature review evaluating relevant road safety literature, industry reports and academic papers on the assessment of CAV safety performance, and a stakeholder engagement exercise to explore concepts of CAV safety assessment with representatives from relevant organizations.

The literature review revealed a lack of specific, quantitative guidance for assessing CAV safety in specific ODDs. This is unsurprising given that, even when broken down into a highly specific ODD, driving is an infinitely variable task. It is therefore extremely challenging to develop precise metrics that account for all potential vehicle factors, environmental factors, traffic factors and so on. Relevant principles from human driving guidance were identified but lack the engineering specificity required for CAVs; mathematical specifications of safe driving that guarantee collision avoidance miss some of the nuances that would encompass what might be considered good driving in the real world. It was therefore suggested that a promising approach would be to mix elements of mathematically safe driving with the system of car control and contextual guidance according to the ODD. The focus would be on outcomes rather than approaches to allow room for innovation.

The stakeholder engagement exercise revealed that existing metrics used for CAV safety are seen as insufficient. However, stakeholders also recognized the challenges in creating objective measurements of highly contextualized behaviours. The importance of capturing near misses in assessing CAV safety performance was also noted as important to improve statistical robustness and sensitivity of evaluations. The ODD was identified as a critical element in managing and evaluating CAV risk.

Based on the literature review and stakeholder engagement tasks, a proposed route to CAV safety benchmarks was offered. Commentary driving is a technique used to train and assess human expert drivers, in which they are required to verbalize relevant information in the driving scene. This is used to determine that they can perceive, prioritize and act effectively when driving. Digital commentary driving (DCD) is proposed as an objective measure of CAV safety performance. DCD does not entail a CAV verbalizing or describing the driving scene in the same way as human drivers, rather, it involves the continuous collection of data from a CAV on its *perceptions*, *decisions*, *reactions* and *feedback* whilst driving. This data effectively probes the "understanding" the CAV has of its environment. Different CAV developers may arrive at this data in different ways (using different sensors, processing techniques, etc.) and their approaches may be highly commercially sensitive. However, since CAV control algorithms must derive this information in order to guide the vehicle, it seems reasonable to collect such information in parallel as the foundational elements of how the CAV determines its future actions. The collection of DCD data is illustrated in Figure 1.

Figure 1 – Diagram illustrating the collection of DCD data from a generic CAV



Analysis of DCD data could then form the basis of benchmarks for CAV safety according to ODDs, with a feedback loop between CAV regulators and CAV developers and manufacturers to promote safety and best practice in CAV operation.

Recommended next steps are for BSI to:

1. Instigate further research with CAV developers to understand the extent to which DCD data can be collected from a CAV, assessing any practical issues in doing so, exploring whether a common standardized DCD data format could be established and determining an appropriate standardization route for this approach (PAS, Flex etc);
2. Work with legal/data experts to resolve potential privacy/security issues in the collection, analysis and storage of detailed CAV data for regulatory purposes;
3. Seek to take forward the DCD approach in government-funded trials where DCD data is analysed from (real or simulated) automated driving compared to similar data from the drives conducted by expert human-driven vehicles and supported by third-party observations to explore how suitable benchmarks for safe driving across the DCD data streams might be set in a variety of ODDs;
4. Examine how DCD data offers synergies with:
    - the BSI PAS 188x series (i.e. supports safe CAV testing per PAS 1881; supports CAV trials data collection per PAS 1882; aligns across ODDs per PAS 1883 etc.);
    - CAV safety performance metrics proposed by other organizations (e.g. SAE Automated Vehicle Safety Consortium, Edge Case Research); and
    - groups developing CAV safety tests to see how DCD can provide a generic feed for their algorithms.
5. Review internal capability, industry appetite and possible standardized approaches for live DCD specifications to be made available in a machine-readable standard format via information management portals.

1 Introduction

Two assumptions are universally accepted around the deployment of automated driving systems (ADSs):

1. they must be at least as safe, but ideally safer than, human drivers; and
2. they will still be involved in serious road incidents.

It has yet to be established what would be considered acceptable in balancing the outcomes of these assumptions, but this is likely to be an essential precursor to widespread deployment of ADSs. In achieving the required safety performance, connected and automated vehicles (CAVs) may adopt behaviours that are considered cautious or risk averse. These may differ from the driving behaviours expected by human drivers, resulting in potential conflicts.

This pre-standardization research project explored approaches to assessing the safety performance of CAVs that would be considered acceptable both in terms of overall safety and from the perspective of other road users. Such assessments were recognized in the context of the operational design domain (ODD) in which the CAV has been designed to operate, as specified within PAS 1883[2]. The focus was on the behaviours and outcomes of complete systems, agnostic to the sensors and algorithms used.

The aim was to deliver a precursor to future standardization of CAV safety assurance, providing regulatory bodies with initial guidance on how CAVs will be developed and deployed in terms of their safety and acceptability, with respect to their use cases of interest. Ultimately, this approach should:

- create minimum performance benchmarks that CAVs must reach to be considered acceptably safe in service (raising the bar);
- provide guidance to CAV developers on the metrics and thresholds to use in developing safe driving behaviours in a given ODD (supporting the industry); and
- provide the public with reassurance about the safety of CAVs as they begin to be deployed on UK roads by enabling the independent evaluation, reporting and regulation of CAV safety performance (raising public confidence).

CAV safety assurance is a prominent research topic. This study aimed to build on and complement current and previous work, anchored by BSI's broader CAV standards programme. This study does not propose nor determine any specific values for what might be deemed suitable or appropriate for individual key performance indicators (KPIs) or over-arching safety metrics but potential approaches for assuring CAV safety performance within a given ODD were investigated.

1.1 Task

The critical task for this study was to explore potential metrics and principles according to the ODD within which a CAV is designed to operate that could be used to provide assurance to the public and raise confidence in the industry over CAV operational safety based on real world performance.

---

[2] PAS 1883:2020, *Operational Design Domain (ODD) taxonomy for an automated driving system (ADS) – Specification.* Available from: www.bsigroup.com/en-GB/CAV/pas-1883/

1.2 Approach

This task was addressed by four organizations with support from BSI:

- ADA (Autonomous Driving Alliance)
- PKSInsights Ltd
- Reed Mobility
- WMG (University of Warwick)

The first phase of the project was to review relevant materials to the topic of CAV safety evaluation to determine where BSI might usefully develop standards to help accelerate progress in the CAV space.

The second phase of the project was to take this proposed approach to key stakeholders and experts in the CAV sector to gain feedback and to help develop the concept further, with the potential to propose new standardized protocols for CAV safety assessment.

1.2.1 Phase 1: Research review

A list of key research topics and materials related to the topic of CAV safety evaluation was collated and distributed across the participating organizations for review, focused on two critical questions:

1. What approach do the authors/originating organization use in relation to evaluating (or specifying) CAV safe driving behaviour?
2. What does this approach tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

By evaluating research materials against these questions, a proposed approach for developing behavioural-led safety benchmarks for CAVs was established.

1.2.2 Phase 2: Stakeholder interviews

Interviewees were selected from a long list of potential participants representing the automotive, simulation, regulatory, research, insurance, communications and CAV developer domains. Time and resource restrictions limited participation to six individuals. Discussions were held in a series of online meetings held between January and March 2021. The format was a 60-minute semi-structured interview under the Chatham House rule. Roles of interviewees and participating organizations are listed in Appendix 1.

Based on the research review and feedback from stakeholders, a proposed research approach for the development of CAV safety benchmarks was created to aid the development of standards to support the CAV sector and transport regulators in safely accelerating the deployment of this technology.

2 Summary of Phase 1 research review

A list of the reviewed materials and comments are provided in Appendix 2.

The reviewed papers did not provide specific, quantitative guidance for automated vehicles against specific ODDs. There is good reason for this: even when broken down into a highly specific ODD, driving is an infinitely variable task. It is extremely challenging to develop precise metrics that account for all potential vehicle factors, environmental factors, traffic factors and so on.

The principles described in the book *Roadcraft* (TSO, 2020) and the concept of "roadmanship" described in Blumenthal et al. (2020) provide the best framework for assessing safe driving. However, they provide no engineering guidance on what safe behaviour means in practice. Conversely, mathematical concepts that deliver fundamentally safe driving behaviours such as "responsibility-sensitive safety" (Shalev-Shwartz, Shammah & Shashua, 2017) and the "safety force field" (Nistér et al., 2019) miss some of the nuances that would encompass what might be considered good driving in the real world. This would include features such as adjusting speed choice in response to the presence of a nearby vulnerable road user, even if their movement could never place them in conflict with the driven vehicle.

Although terminology differs slightly, *Roadcraft* (TSO, 2020), the *Safety First For Automated Driving* white paper (Aptiv et al., 2019) and other research papers in the field of robotics and human skilled performance characterize behaviour in three steps:

1. *perception* (status, location, objects, applicable rules etc.);
2. *decision* (how to prioritize and achieve desired objectives in the context of current perceptions); and
3. *reaction* (actions implemented to achieve desired objectives).

This reveals a potentially useful framework with which to categorize data captured from a CAV in the interests of assessing its safety performance.

The proposed approach therefore is to mix elements of mathematically safe driving [like responsibility-sensitive safety (RSS)] with the system of car control and contextual guidance (like *Roadcraft*). This would entail developing principles for a structured approach to collecting data across perception, decision and reaction categories that enable tailoring of metrics according to ODDs. The focus would be on outcomes rather than approaches to allow room for innovation.

3 Summary of Phase 2 stakeholder engagement

Stakeholders agreed that the widely known measures of CAV operations (in particular collisions, miles driven and number of disengagements, e.g. California Department of Motor Vehicles, 2020) are insufficient to provide a realistic assessment of CAV safety performance. Many additional metrics were suggested including:

- vehicle kinematics [i.e. measures of vehicle dynamic performance such as velocity, acceleration, jerk (all in multiple axes) etc.];
- unexpected stops (i.e. false positives: the CAV stops when there is no discernible reason why the ADS acts to bring the vehicle to a halt);
- infractions (i.e. CAV passes traffic light on red or exceeds the speed limit); and
- alignment with behaviours that could be considered *Roadcraft* [i.e. as described in the police drivers' handbook, *Roadcraft* (TSO, 2020)].

Such metrics are listed here in ascending order of difficulty in terms of the challenge to developers in providing data on safety performance of their CAVs. Kinematic measures can be obtained directly from sensors. Stops are also simple to detect but context is required to determine whether they are truly unexpected or a reasonable action in relation to the driving scenario encountered. Infractions require more knowledge about the context in which a behaviour was observed and if, for example, a CAV failed to detect a traffic light showing a red signal, this would not be recorded as an infraction. Alignment with defined behaviours in *Roadcraft* bring more subjectivity to observation. For example, in describing “The system of car control”, *Roadcraft* states that a driver should adopt a position to “negotiate hazard(s) safely and smoothly” – with similar guidance over speed, gear and acceleration choices. These make intuitive sense to a human driver but may be ambiguous to a CAV developer and highly context-specific. For this reason, stakeholders suggested the need for digital road rules, whereby these ambiguities are addressed (recognizing that there may be situations in which such rules need flexibility in the interests of safety).

While *Roadcraft* offers some useful insights into safe driving by humans, care was advised by stakeholders over the temptation to compare automated driving directly with human driving behaviours. This was raised firstly, because the actions of the CAV in the context of its use case may be entirely adequate for this specific purpose but different to the actions that a human driver may take, and secondly, because human driving does not always represent the highest standard of driving behaviour (e.g. speed compliance, passing distance to cyclists etc.).

One CAV safety metric that requires clearer characterization is the term “near miss”. With collisions representing the tail of the distribution of driving behaviours, there is value in capturing crash-relevant events to provide some statistical insights into the likelihood of crashes (i.e. even if no crashes are observed, we can draw some conclusions on CAV safety performance based on the number of near miss events observed for CAV type A vs. CAV type B). However, as with infractions, it may be challenging for a CAV to recognize that a near miss has occurred. It was therefore suggested that third-party observation (expert observers within the vehicle or monitoring from other vehicles/infrastructure) may help to identify near misses.

The importance of the ODD was highlighted by all stakeholders as means by which critical risks may be identified and mitigated ahead of CAV deployment. This highlights that a standardized approach to assessing CAV safety risks aligned to ODD descriptions may be useful for CAV developers in mitigating risk and proving safety performance of their vehicles.

In relation to the two use cases discussed (Urban Robotaxi and Highway Pilot), stakeholders indicated that the potential complexity of urban operations (two-way traffic, complex junctions, pedestrians, crossings, etc.) vs. the relative simplicity of highway driving (unidirectional traffic flow, absence of more vulnerable road users, fewer/less complex junctions, etc.) would likely dictate that there would be more safety metrics than for highway CAV operations.

Stakeholders recognized the importance of data sharing. For safety performance metrics to be useful there needs to be a standardized approach to assessment so that sharing of data enables objective assessment of CAV safety. Furthermore, one stakeholder suggested that CAV safety metrics could be captured by a separate “black box” in the vehicle to capture a standardized, independent assessment of CAV activity. This could build on many years’ experience in the use of event data recorders (EDRs) in vehicles, though would require more information to reflect the complexity of CAV operations.

Although collisions remain the ultimate metric for safety, the most important dimensions to assessing CAV safety performance according to stakeholders were error margins and uncertainty over perceptions and actions, both of which will be highly conditional upon the situation.

4 Proposed route to CAV safety benchmarks

4.1 What makes a good driver?

Although a statistically unlikely number of human drivers tend to believe they are "good" drivers (Finn and Bragg, 1986; McCormick et al., 1986; Marottoli and Richardson, 1998; McKenna, Stanier and Lewis, 1991; Harré et al., 2005), the definition of a "good" driver is elusive. In UK law (Road Traffic Act, 1988), the definitions of both "careless driving" and "dangerous driving" each refer to situations in which the exhibited behaviour of the driver in question "falls below what would be expected of a competent and careful driver". For drivers on British roads, those expectations are captured in the various rules and guidance set out in the Highway Code (Driver and Vehicle Standards Agency, 2015).

The CertiCAV project is led by the Connected Places Catapult on behalf of the UK Department for Transport and aims to create a framework of safety criteria for CAVs. The project defined nine principles that they suggest constitute as good driving:

1. Do not cause collisions.
2. Avoid collisions, even when not the cause.
3. Provide occupant safety/comfort.
4. Follow traffic rules.
5. Provide reasonable safety margins.
6. Follow recommended driving practice.
7. Facilitate established driving conventions.
8. Behave considerately to other road users.
9. Do not unreasonably obstruct movement of traffic.

[Source: CertiCAV project, Connected Places Catapult, 2021]

However, both the Highway Code and the CertiCAV principles leave considerable room for interpretation. This is acceptable in the context of human drivers where instincts for safety, for cooperation and for predicting others' behaviour can finesse any ambiguous, qualitative guidance. However, without those instincts, it is difficult for a programmer to create code that can suitably interpret complex situations and determine a safe course of action. For example, it is relatively easy for a human driver to understand how considerate behaviour towards other road users might play out in a wide range of situations but much harder for these behaviours to be codified for a machine.

Before being permitted to drive on public roads alone, human drivers are typically assessed by a formal driving test that assesses their competence against a range of criteria. To drive a car in the UK, this includes a theory test (multiple choice questions and hazard perception test) and a practical test (approximately 70 minutes and includes a variety of specific manoeuvres, directed driving in general traffic and independent driving). However, collision risk in the first 1 000 miles after licence acquisition is still likely to be the highest it will ever be throughout a driver's lifetime (Helman, Grayson & Parkes, 2010). Achieving an equivalent standard for CAVs seems insufficient. While it is likely there will be some form of pre-deployment certification tests for a CAV, the data produced by a CAV could be used for continuous evaluation of performance. Insurers selling motor insurance policies based on telematics systems attempt to do this in relatively simplistic ways with kinematic triggers and driving time restrictions monitored through the driver's smartphone or a vehicle-mounted data recording and communication device (often referred to as a "black box" – see RoSPA, 2013).

IAM Roadsmart suggested an automated vehicle should perform at least as well as an advanced driver in any given situation. This assertion triggered exploration of the ways in which advanced drivers are trained and assessed for their potential relevance in CAV safety evaluation. Further guidance on what constitutes a good driver is provided by *Roadcraft* (TSO, 2020), subtitled *The Police Driver's Handbook*. They list the following characteristics as being the qualities of a safe and competent driver:

- Critical and honest self-awareness and understanding of your personal characteristics, attitude and behaviour, which are necessary for safe driving.
- Taking action to keep identified risks to a minimum.
- Awareness of your own limitations and those of the vehicle and the road.
- Awareness of the risk inherent in particular road and traffic situations.
- Concentration and good observation.
- Continuously matching the vehicle's direction and speed to the changing conditions.
- Skillful use of vehicle controls.

[Source: *Roadcraft*, TSO, 2020]

Roadcraft describes how:

> *"Safer driving depends on systematically using the information you gather from observation to plan your driving actions".*

[Source: *Roadcraft*, TSO, 2020]

Illustrating with an example, *Roadcraft* describes how a driver must:

a) observe current and future hazards;
b) anticipate how the scenario is likely to evolve;
c) prioritize the most significant hazard(s);
d) decide what to do in response; and
e) act in response to the hazard(s).

This process is also characterized as a continuous loop in which:

- a driver receives *inputs* (information from the senses and from stored experiences);
- are used for *decision making* [planning appropriate actions in the context of the driving situation (and how it may develop) and objective(s)];
- leading to *outputs* (taking the required actions); and
- *feedback* (new information received as a result of taking the intended actions) – which produces new *inputs*.

Similarly, in the proposal for UN regulations on automated lane keeping systems (UNECE, 2020), the skilled human performance model for driving is described as having three phases – *Perception*, *Decision* and *Reaction*. These formulations of information processing seem to offer intuitive relevance for considering how a CAV may detect, process and act upon data collected from its sensor systems in order to deliver safer driving and create a framework for comparison between the safety performance of human drivers and CAVs.

Also cited in *Roadcraft* are the three core competences for all drivers, derived from the *Goals for Driver Education framework* (Hatakka et al., 2002):

- the knowledge and skills to drive safely;
- an understanding of the factors that increase your risk of a collision; and
- the ability to accurately assess your driving behaviour.

[Source: *Roadcraft*, TSO, 2020]

Roadcraft recommends that police and emergency service drivers develop these three competences to the highest possible standard. One of the ways in which advanced drivers, including police/emergency service drivers are trained and assessed is through "commentary driving". This is a technique whereby the driver describes in real-time:

- what situation relevant items they can see;
- what static, dynamic and environmental features they are considering; and
- what actions they may be planning to manage risk and continue safely.

Coupé et al. (2019) identified that human speech is typically limited to an information rate of approximately 39 bits per second. This information bottleneck means that a human driver may therefore be unable to describe all the relevant information present in a complex driving environment. The items the driver chooses to prioritize in their commentary are therefore also a significant indicator of hazard awareness in their training and assessment. Several studies have noted the beneficial effects of commentary driving on drivers' hazard perception performance and safety (e.g., Mills et al., 1998; McKenna et al., 2006; Isler et al., 2009, Crundall et al., 2010).

The commentary drive allows an instructor/examiner to understand:

- the (prioritized) items the driver is seeing;
- how they imagine the situation will unfold;
- the features of the situation or vehicle are influencing their decision making;
- the actions they are planning in response.

Each of the items represents pertinent information that should be obtainable from a CAV to understand whether it can be considered to be driving safely. An instructor guiding the trainee driver delivering a commentary provides a source of feedback and tuition to the driver, supporting improvements in their future performance.

However, unlike a human driver, data recording systems for a CAV are not constrained by the same speech bit rate so a richer, continuous data stream could be available for a CAV (while recognizing that the way in which hazards are prioritized by a CAV is still a topic of critical importance).

4.2 What makes a good mobile robot?

Professor Paul Newman, director of the Oxford Robotics Institute and co-founder of the automated vehicle software company, Oxbotica, often references three key questions that drive his research (e.g. Ingenia, 2015; Oxbotica, 2017; Volvo, 2019). Newman states that a mobile robot must answer the following questions:

1. Where am I?
2. What is around me (and what are they doing)?
3. What should I do next?

This structure echoes other conceptualizations of how systems (including humans) manage behaviour in complex environments, simplified into three steps:

- *Sense* important information about the environment (e.g. detecting and identifying pedestrians);
- *Plan* how to respond to achieve the objective (e.g. drive safely from origin to destination) in the context of the sensed information (e.g. pedestrian about to step into the road); and
- *Act* to deliver the plan (e.g. apply the brakes to avoid the pedestrian).

This framework is applied in the automotive context in the *Safety First For Automated Driving paper* (Aptiv, Audi et al., 2019) and aligns with the descriptions in *Roadcraft* and the skilled human performance model for how human drivers process and respond to sensory information – actions are determined by a continuous loop of perceptions, decisions, actions and feedback.

Essentially, safe driving depends on a human or an ADS being able to answer Newman's questions successfully. This approach has parallels with commentary driving, where the human driver essentially gives a distillation of their answers to these questions verbally to an instructor (whose assessment of their driving and commentary performance acts as a feedback loop for evaluation or training). The commentary process is therefore a technique that seeks to access the underlying cognitive information that a driver is using to inform their actions.

For a mobile robot, developers have to determine from the ground up how their system is going to answer these questions, based on the software and hardware tools at their disposal. Consequently, it should be possible to access the variables, parameters and instructions that critically determine how the system is operating. However, if aspects of the robotic operation are based on machine learning, some of the information processing may be inaccessible. Rudin (2019) makes a strong case as to why high stakes decision making models need to be transparent and accountable rather than restricted to more opaque machine learning approaches. With the operation of motor vehicles at speed in public environments presenting the potential for loss of life, automated driving can certainly be considered to be a high stakes activity; therefore, CAVs operating based on machine learning models still need to be transparent to enable examination of the factors that influence their behaviour.

This point is reinforced in the European Commission's proposed regulation of systems using artificial intelligence (European Commission, 2021). The proposal classifies artificial intelligence (AI) systems used in critical infrastructures for transport (e.g. CAVs) as "high-risk" and that providers of high-risk AI systems should keep logs for monitoring system operation to specified standards (Article 12) for a required period (Article 20) and submit those logs to competent national authority in response to a reasonable request for access (Article 23). Similarly, the Chinese Ministry of Industry Information Technology (MIIT) released draft guidance on the management of intelligent connected automobiles, which states that CAVs must have event data logging and data storage functions to include (at least) the operational status of the automated driving system, driver status, driving environment information, and vehicle control information, and shall meet relevant performance and safety requirements to ensure the integrity of the data logged by the device in the event of an accident.

In recommendations to the European Commission on the ethics of CAVs, Bonnefon et al. (2020) noted that continuous monitoring of CAV safety and the traceability of pre-crash information may be critical in considering CAVs to be operating safely and ethically. The ITU (International Telecommunications Union) has established a dedicated focus group (FG-AI4AD) for the behavioural evaluation of automated driving. In a presentation to the UNECE working party 1 (WP.1 – Global Forum for Road Traffic Safety) and working party 29 (WP.29 – World Forum for Harmonization of Vehicle Regulations), FG-AI4AD advocated for the adoption of post-deployment behavioural monitoring of CAVs using "leading" measures – evaluations of safety performance that enable robust statistical analysis of the likelihood of collisions (see Balcombe, 2020).

Having a standardized, comparable means by which to access the information that a CAV has used to guide its actions and the driving behaviours adopted as a result seems to fulfil these recommendations.

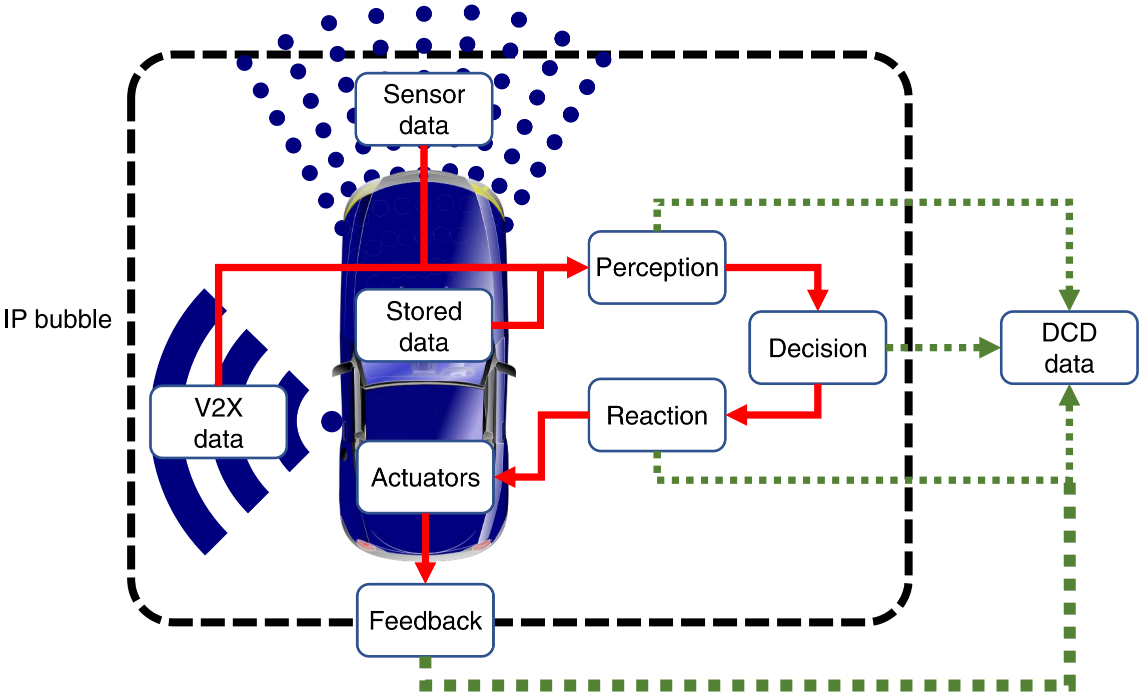
4.3 Digital commentary driving

To align approaches to assessing the qualities of a good driver and those of a good mobile robot, **digital commentary driving** (DCD) is proposed. DCD does not require that a CAV should be providing any kind of narrative description of its journey. Instead, DCD comprises a standardized set of data parameters that support analysis of CAV driving competence. By accessing the perceptions, decisions, reactions and feedback that constitute automated driving by a CAV, DCD seeks to provide insights into whether the CAV is operating safely. With the goal of CAVs being safer drivers than humans, DCD enables any incident or near miss to be analysed from the CAVs point of view in non-expert terms, comparable between different vehicle types.

Unlike a human driver, these parameters are not constrained by the limited bandwidth of speech so more data can be captured continuously throughout every drive completed by a CAV. Furthermore, the same DCD data can be captured in simulation for comparison and validation against real world DCD data. However, without the human ability to contextualize what is being seen, analytical tools will be required to determine whether a CAV is behaving appropriately based on the data collected – and therefore deliver feedback on CAV safety performance.

Figure 1 illustrates how DCD data would be collected from a generic CAV.

Figure 1 – Diagram illustrating the collection of DCD data from a generic CAV



The red arrows in the diagram show the flow of information within the ADS to enable automated driving. The data collected and the information processing remains proprietary to the ADS developer, sitting within an IP "bubble". Data flows from internal and external sources to provide a continuous perception of the outside world. From this perception, decisions are made about how to proceed, leading to a reaction implemented by vehicle actuators and feedback on whether the intended reaction has been achieved. Feedback sits on the IP bubble since, while some elements of feedback will be internally registered, actions of the vehicle are externally observable (i.e. its driving behaviour) and therefore not protectable. The dotted green lines show the collection of DCD data as a specific, standardized extract of information from the perception, decision, reaction and feedback processes in the CAV. Developers may differ significantly in how data flows between these boxes and the processing within them but the outputs to DCD data should be the same. DCD is therefore agnostic of sensors, processing and IP used to perceive, plan and implement automated driving. This consistent, standardized data is the only information to permeate the IP bubble for the external assessment of CAV safety by regulators.

The following table provides a (non-exhaustive) list of potential data streams to be captured during DCD, each categorized as perception, decision, reaction or feedback.

Table 1 – Proposed sample of potential digital commentary driving (DCD) data streams, categorized by their contribution to *perception*, *decision*, *reaction* or *feedback* processes

Perception	**A CAV should report:**
Location	Information about its location, whether derived from external sources (e.g. global navigation satellite system) or localization against internal data stores (e.g., high-definition map). DCD data should include the CAV's location estimate and a confidence level about the accuracy of this estimate.
Vehicle status	Current status of actuators (e.g. steering direction, accelerator/brake actuation). Whether all systems involved in automated driving (e.g. sensors, processors, actuators) are operating as expected, along with a record of relevant software versions operating.
ODD awareness	Whether (or not) it is within its ODD and if there are any parameters that suggest that it might exceed its ODD for any reason (e.g., nearing the edge of a geofence, visibility conditions reaching a critical level etc.).
Kinematics	Velocity and acceleration (both linear and rotational in multiple axes). Individual wheel speeds should also be recorded to determine any under- or over-rotation.
Object detection	When each individual object (e.g. traffic light, vehicle, pedestrian, cyclist, horse etc.) was detected, when and how it was classified and how far away it is perceived to be. Each of these assessments should be provided with confidence intervals.
Object predictions	Estimates of how dynamic (or potentially dynamic) objects are moving (velocity) and how they are predicted to move, again with confidence intervals on those estimates.
Local environment	Road rules that are applicable at its current location (e.g. speed limit, line crossing etc.) and any relevant environmental conditions that may influence its behaviour (e.g. risk of black ice). The mode and strength of its network connection (e.g. 5 G, 250 Mbps) where relevant (and if required for safe driving).
Decision	**A CAV should report:**
Goal prioritization	How its goals were prioritized in the light of the current driving situation (e.g. continue at current speed to reach destination at predicted time vs. slow down to mitigate collision risk).
Object prioritization	How objects were prioritized in terms of criteria for avoidance (i.e. the extent to which each is considered a hazard).
Reaction	**A CAV should report:**
Path	Future planned trajectory along with a confidence interval on the accuracy of that path.
Target speed	Speed that it is aiming to achieve based on its current perception of its environment.
Feedback	**A CAV should report:**
Compliance	Whether behaviour contravenes any of the applicable road rules (e.g. exceeding the posted limit) and report on what basis the decision to do so was made.
Closest approach	Minimum distance observed when passing objects (and speed at point of minimum distance).

Automated driving depends on the collection and processing of large volumes of data from a variety of sources. However, DCD data represents an extract of outcomes of the processing of this data and would therefore be a tiny fraction of the size of the total data used in automated driving (DCD data likely to be in the order of a megabyte per hour of driving).

5 Discussion and recommendations

The ability to access live data on the relative safety performance of an automated vehicle is a highly prized objective with many organizations working to address this topic. This seems to be a vital step in gaining assurance about the acceptable performance of a CAV. The proposed DCD approach is a starting point towards a process that regulators, manufacturers, operators, users and the general public can use to build confidence that CAVs operate in ways that are sufficiently safe and ethical. If the DCD process were to become established, it would enable the development of a suite of key performance indicators, tailored per ODD for the ongoing assessment of CAV safety performance. This would give regulators tools with which to assess and compare relative CAV safety performance as well as providing vital information to support determination of liability and establishing contributory factors in the event of a collision involving a CAV.

The demands of the DCD approach on the CAV developers/operators are significant in terms of the requirement to produce and store data. A balance must be struck between being able to operate CAVs that meet industry and public expectations of safety performance and ensuring that this does not create demands that unduly delay access to the potential benefits of CAVs. A standards-based approach might offer benefit in that regard, allowing rapid iterative development with critical involvement from the public, private and academic sectors.

The development of CAV technologies is a competitive space with developers commensurately protective over their intellectual property (IP) in CAV operation. Probing the data produced by CAVs as part of DCD could cause developers some discomfort over the potential for IP to be revealed. However, the DCD data has been selected not only for its suitability in determining CAV safety performance but also on the principle that the requested data streams must be calculated, estimated or derived in the process of automated vehicle operation. DCD makes no requirements for developers to reveal anything about *how* the requested information is derived, only that it should be made continuously accessible for any drive. Onboard DCD data collection could be achieved through a (real or virtually) segregated recording device with no access to any onboard data processing activities. Furthermore, a regulator using DCD data would not share results from individual companies but may provide aggregated, anonymized feedback to CAV operators to help improve safety performance. These protections mitigate the risk of IP loss. This concept was illustrated in Figure 1, which highlights the flow of information within a generic CAV and the extraction of DCD data from the process.

An important aspect of DCD is that it would provide regulators with the evidence to make more sensitive judgements about the safety performance of CAVs. In the absence of DCD, an operator may be required to withdraw their CAVs from commercial service following a minor incident with significant lost revenues as a result. The presence of DCD data would support detailed analysis as to why the incident took place so that regulators could apply more proportionate limitations on CAV operations, for example, if the data suggested the incident was caused by perception failures in low light conditions, the operator could be required to modify the ODD for their CAVs to prevent operation at night. However, commercial services could continue in brighter conditions, allowing continued revenue generation. Conversely, this evidence would also provide a regulator with stronger evidence to enforce withdrawal of service if DCD reveals a more fundamental issue.

A key finding from the research review was the limited extent to which ODDs were exploited for the evaluation of CAV safety. However, the stakeholder engagement process revealed considerable differences in expectations of performance and key metrics in the two SAE L4[3] CAV scenarios discussed (Urban Robotaxi vs. Highway Pilot). DCD data collected from CAVs operating in different ODDs will help to generate benchmarks and ranges of acceptable performance that can be used by developers to ensure that their vehicles behave

---

[3] Society of Automotive Engineers (SAE) (2018). J3016, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Available from: www.sae.org/standards/content/j3016_201806/

safely and by regulators to ensure deployed CAVs are meeting industry and societal expectations over safety performance (recognizing that these metrics may evolve over time).

Developers may be concerned that the real-time production of DCD data may interfere with critical computer processing activities in ways that harm the core functionality of CAV operation. However, DCD data need not be produced in real-time. However, it seems appropriate for CAV operators to derive DCD data offline, provided that this can be achieved for any part of any drive based on their own data records. If DCD data is generated offline, it also critical that this process is transparent such that the data is a true reflection of the perceptions and actions of the CAV and not manipulated for any reason.

Similarly, developers may be concerned that DCD data may result in inordinate data storage requirements. However, all of the data streams captured are simple text outputs that can be readily stored and compressed to minimal file sizes, certainly in comparison to video and lidar data files. A further issue that could expand DCD data file size is frequency of recording. Since DCD data is primarily concerned with hazard awareness and vehicle behaviour, a relatively low frequency of data collection can be applied, with 10 Hz as a possible starting point for discussion. This could be accompanied by higher frequency recording based on kinematic or event-based triggers to permit more in-depth analysis of safety-critical events, along the lines used by event data recorders (UNECE, 2021).

For a human commentary drive, the instructor can see the driving environment and assess whether the driver is evaluating the scene correctly. A real challenge for the DCD approach is the lack of objective "ground truth" information for comparison. In the event of a serious incident, DCD data could be combined with any recorded video data from the CAV (and any supplementary materials from roadside cameras other vehicles etc.) to provide further insights into the cause of the collision. However, if a CAV were to have a near miss with a pedestrian that it had failed to detect, no information about that near miss would be available in the DCD data. As noted earlier, the DCD format could also be used in CAV simulation where ground truth data about all objects in a scene is available. Simulations, along with real-time comparative assessment by human expert observers and other third-party observations (from roadside cameras, other vehicles etc.), could be used to verify DCD data and to help determine (and update) suitable metrics before they are applied in the real world.

A key benefit of DCD is that it can be used for CAV assessment even while the ADS is not in active control (offline DCD). In this offline DCD mode, the ADS performance of perception and decision modules in completing the object and event detection, the hazard identification and the proposed plan for mitigating the hazards can be made while a human driver executes the driving task. This has significant safety benefits for early trials, initial assessments or as a phased approach to new software releases.

A tangential benefit of the use of offline DCD is that, for a deployed CAV under human control, the ADS data collection would provide an objective technique by which to assess the relative safety performance of human drivers against the same safety benchmarks used to assess for safe automated driving.

Finally, discussions with BSI revealed an intriguing possibility whereby a potential future live DCD standard could be made available via API access on widely used information management portals such as GitHub. This could help CAV developers and operators ensure that they are working with the most up to date version of DCD specification.

Recommended next steps – BSI should:

1. Instigate further research with CAV developers to understand the extent to which DCD data can be collected from a CAV, assessing any practical issues in doing so, exploring whether a common standardized DCD data format could be established and determining an appropriate standardization route for this approach (PAS, Flex etc);

2. Work with legal / data experts to resolve potential privacy/security issues in the collection, analysis and storage of detailed CAV data for regulatory purposes;
3. Seek to take forward the DCD approach in government-funded trials where DCD data is analysed from (real or simulated) automated driving compared to similar data from the drives conducted by expert human-driven vehicles and supported by third-party observations to explore how suitable benchmarks for safe driving across the DCD data streams might be set in a variety of ODDs;
4. Examine how DCD data offers synergies with:
    - the BSI PAS 188x series (i.e. supports safe CAV testing per PAS 1881; supports CAV trials data collection per PAS 1882; aligns across ODDs per PAS 1883 etc.);
    - CAV safety performance metrics proposed by other organizations (e.g. SAE Automated Vehicle Safety Consortium); and
    - groups developing CAV safety tests to see how DCD can provide a generic feed for their algorithms; and
5. Review internal capability, industry appetite and possible standardized approaches for live DCD specifications to be made available in a machine-readable standard format via information management portals.

6 Appendix 1. Phase 2: Stakeholder interviews

Table 2 lists the roles and organizations of individuals involved in the interviews as part of the stakeholder engagement process.

Table 2 – Participants in stakeholder interviews

Interviewee role	Representing
Head of Policy & Research	IAM Roadsmart
Lead Lawyer	Law Commission
Senior Scientist	Met Office
Senior Engineer	Nissan
Director of Research	Thatcham
Senior Safety and Risk Consultant	TRL

7 Appendix 2. Phase 1: Research review

Critical materials were collated for review by the project team with two specific questions addressed for each item:

1. What approach is used in relation to evaluating (or specifying) CAV safe driving behaviour?
2. What does it tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

7.1 BSI (2020), PAS 188x

What approach is used in relation to evaluating (or specifying) CAV safe driving behaviour?

The published BSI PAS documents provide useful insights into ways in which CAV safety might be evaluated. PAS 1880 relates to the assessment of the safety of control systems for CAVs. While it does not provide deep insights into driving behaviours that a CAV should adopt to achieve safety, it does provide some general safety and system architecture requirements associated with sensor operation that should be met to manage risk when operating a CAV.

PAS 1881 relates to assurance for CAV trials and the creation of a suitable safety case. However, many of the requirements for the safety case are potentially relevant to CAV commercial operation. For instance, a requirement to demonstrate that data is adequately preserved and processes to identify and report near misses.

PAS 1882 relates to data collection and management for CAV trials. This document is insightful as it contains a set of data streams that CAV trialling organizations should collect to provide assurance about safe CAV operation.

PAS 1883 relates to the definition of the operational design domain and is therefore directly of interest to this research.

What does it tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

Identified gap: The development of standards for CAV safety benchmarking was recognized as a gap in the standards landscape so it is no surprise that the BSI PAS 188x series does not meet this requirement. However, it does provide a wealth of useful information about the ways in which CAV trialling organizations should act to provide assurance over safety performance, with resonance over how those principles might apply in CAVs when commercially deployed.

7.2 Koopman et al. (2019), UL 4600 and SPIs

What approach is used in relation to evaluating (or specifying) CAV safe driving behaviour?

UL 4600 is designed specifically for CAVs and is intended to extend e.g. ISO 26262, ISO/PAS 21448 to provide more appropriate standards, designed to cope with:

- AI-driving behaviours and associated unpredictability;
- the implications of having no human driver;
- changes to the environment that require updating in digital maps;

- pace of technological change; and
- ensuring safety in an immature technological environment.

UL 4600 is focused on creating a suitable safety case for CAV operation. Behaviours are goal-driven, contextually described and technology agnostic, with rapid feedback loops for collecting and sharing information about uncertainties.

Safety cases are provided at the component level as "plug-ins" to the overall system safety case.

Requires CAV developer to provide evidence that specific safety arguments made in the safety case are addressed; monitored by safety performance indicators (SPIs).

What does it tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

Highly detailed safety case approach with standardized approach to safety case and the valuable approach to safety assurance through the use of SPIs. However, SPIs are developed by a testing organization to address the specific demands of the scenarios that they expect their vehicles to encounter.

Identified gap: No guidance is provided to a testing organization about the sorts of acceptable performance that might be appropriate for the ODD of their vehicles. This is a gap that BSI research could attempt to address.

7.3 Nistér, Lee, Ng, Wang (2019), *An introduction to the safety force field* (nVidia)

What approach is used in relation to evaluating (or specifying) CAV safe driving behaviour?

The safety force field is defined by a general theory of safety at the obstacle avoidance level. The theory leads to a computational mechanism for mapping world perception into constraints on control that, if obeyed, prevents all collisions. This approach is based on the premise that systems that are compliant with all international traffic rules are desirable but the foundation of safety should not depend on that long tail of rules. The core of the concept is defined around dynamic objects being able to perform a "safety procedure" which effectively means slowing to a safe stop:

> *"All actors are required to perform their safety procedure (or better) before and whenever the trajectory resulting from their safety procedure intersects with that of another actor. If all actors do what they are required, no collisions can occur".*

There follows a detailed mathematical exposition of the core concept based around 21 definitions and 10 lemmas.

What does it tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

The authors provide a highly detailed mathematical description of how automated vehicles can operate in a dynamic environment while avoiding collisions with any other static or dynamic objects.

Identified gap: However, it does not attempt to specify or determine how specific properties of the environment or other objects might influence the behaviour of the driven vehicle. For example, an automated vehicle passing a stationary pedestrian at 1 m lateral clearance would be displaying equally valid behaviour at 3 ms^{-1} as at 30 ms^{-1}; however, the experience for the pedestrian would be markedly different in each case. The opportunity therefore exists for research to explore how the safety force field approach would be adapted according to ODD and specific properties of static and dynamic objects that should cause an AV to modify its behaviour.

7.4 Waymo (2020), *Safety case and public road safety performance data*

What approach is used in relation to evaluating (or specifying) CAV safe driving behaviour?

Waymo is acknowledged to be a global leader in roadgoing CAV operations with automated vehicles, offering commercial ride-hailing services to passengers without safety drivers in Arizona, USA; the first company to achieve this feat. Their safety framework describes how they develop, evaluate and deploy CAVs in the real world using a portfolio of methodologies to embed safety in their hardware ("performant, secure and robust"), behaviour ("safe and responsible") and operations ("safely deployed and operated") – all grounded in a broader safety governance structure (Waymo Safety Board, safety focused internal culture and processes).

Decisions about how CAVs should drive are assessed using three methodologies:

1. hazard analysis;
2. scenario-based verification (includes collision avoidance testing in simulation and closed track trials); and
3. simulation (millions of miles tested using high quality human driving as the benchmark; includes examination of scenarios where a human driver took over control to understand whether automated systems would have responded appropriately).

Performance within these methodologies are assessed using three main metrics: collision avoidance, trip completion and rule compliance.

Their public road safety performance data relates to 6.1 million miles of driving in Phoenix, Arizona, including 65 000 miles of automated driving without a safety driver. These data showed 47 contact events, none of which resulted in serious injury and for almost all, rule violations or other errors by a human driver was involved.

What does it tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

The Waymo documentation is an excellent exposition of how they have structured their approach to safety.

Identified gap: However, while they show how their CAVs have performed relative to the collision frequency of human drivers, details are not provided on how they have determined the acceptable driving behaviours that their CAVs adopt within their ODD. This is perhaps not surprising as they may see this as being commercially sensitive information.

7.5 TSO (2020), *Roadcraft*

What approach is used in relation to evaluating (or specifying) CAV safe driving behaviour?

This book is marketed as "The Police Driver's Handbook" and is intended as companion to the training courses undertaken by emergency services drivers but also supports civilian drivers in seeking improve their driving skill aiming to:

> *"...help you become a better driver by increasing your awareness of all the factors that affect your driving – your own capabilities, the characteristics of your vehicle, and the road and traffic conditions".*

CAV developers can seek inspiration from the ways in which Roadcraft enables human drivers to improve their ability to assess risk on the road and moderate their driving behaviours accordingly. The fundamental basis of Roadcraft is the 'system of car control'. This guides the driver in how to:

- use the sequence of position, speed, gear and acceleration to manage the vehicle safely in response to hazards;
- continuously take, use and give information regarding the environment and the actual or possible behaviours of other road users.

Examples are given as to how the system of car control can be applied to tackle common manoeuvres safely. Similarly, examples are given on how to recognize potentially hazardous situations and exercise caution. Issues around the perception of other roads users' signals and how they may be ambiguous are very relevant for CAVs in predicting likely future actions. Some of the details about gear control have some relevance for CAVs but as we move towards EV powertrains the value of this information will diminish. Useful guidance is given on road position and overtaking.

What does it tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

Many of the terms within Roadcraft are framed from the perspective of a human driver but have relevance for CAVs. Like a human driver, a CAV must continuously observe its environment and use contextual information to decide upon and enact an appropriate course of action. In this way, much of the guidance in Roadcraft is highly relevant to the safe CAV driving behaviour, especially because CAVs are likely to be sharing the roads with human vehicles for the foreseeable future so will need to behave in a predictable and intuitive manner for other road users.

Identified gap: Since driving behaviour is highly dependent on situational factors, no specific values are given on issues such as following distances or passing clearances; however general principles are provided to guide behaviour.

7.6 The Law Commission (2019-2020), Consultation papers

What approach is used in relation to evaluating (or specifying) CAV safe driving behaviour?

The Centre for Connected and Autonomous Vehicles (CCAV) asked the Law Commission of England and Wales and the Scottish Law Commission to undertake a far-reaching review of the legal framework for automated vehicles, and their use as part of public transport networks and on-demand passenger services. Much of this work has taken the form of three open consultation processes; in December 2020, two have been completed, the third and final consultation is imminent. As CAV regulation is a topic with broad public and multi-industry relevance, these consultations have drawn responses from many organizations. Where organizations have not specified otherwise, consultation responses have been published in full on the Law Commission's website.

Some of the questions within the consultations have relevance to how safe driving behaviour might be evaluated. In Consultation 1, questions 38-43 asks about CAV behaviours and compliance with various rules of the road.

What does it tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

Unsurprisingly, many organizations felt CAVs should be law abiding. However, some differences emerge in discussing the situations where CAVs might be permitted to contravene the rules of the road. For example, whether it would be permissible for CAVs to exceed the speed limit for traffic flow reasons. A key principle identified was that regulations should be enforceable and should focus on outcomes rather than approaches. This means that the desired effect is achieved without constraining how it is delivered, leaving room for innovation in approach.

7.7 Automated Vehicle Safety Consortium (2020)

What approach is used in relation to evaluating (or specifying) CAV safe driving behaviour?

AVSC is an SAE, Ford, GM, Daimler, Lyft, Honda and VW consortium looking to establish safety principles, common terminology and best safety practices. There is much focus on event data recording, with acknowledgment that "event" needs some definition but suggests this might involve significant velocity changes or non-reversible elements such as airbag deployment. A near miss probably does not constitute an event. Event data includes:

- Vehicle control – vehicle state and requested state.
- Saliency – what the vehicle thought was important.
- Sensing and General Parameters – what the vehicle saw. Refers to SAE J3197 so thinking likely to be raw data rather than perceived objects of type and expected behaviour.
- Data should be recorded at 0.2 Hz before an event and higher if needed for event analysis.

The actual question of what constitutes safe appears to be the OEMs' private task; AVSCs do not list safety KPIs rather the data that would be useful in the investigation of an incident.

What does it tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

AVSC have published papers[4] on:

- *Best Practice for First Responder Interactions with Fleet-Managed Automated Driving System Dedicated Vehicles (ADS-DVs)*;
- *Best Practice for Data Collection for Automated Driving System Dedicated Vehicles to Support Event Analysis*;
- *Best Practice for Passenger-Initiated Emergency Trip Interruption*;
- *Best Practice for Describing an Operational Design Domain: Conceptual Framework and Lexicon*; and
- *Best Practice for safety operator selection, training and oversight procedures for automated vehicles under test.*

Identified gap: No benchmarks or thresholds that might be used to determine how safe an ADS is have yet been established but it seems these may be coming in 2021.

---

[4] A full list of AVSC publications is available from: https://avsc.sae-itc.org/best-practices-5471WV-45892W7.html?respondentID=29057494#our-work

7.8 Mobileye (2017), *Responsibility-Sensitive Safety (RSS)*

What approach is used in relation to evaluating (or specifying) CAV safe driving behaviour?

The concept of RSS is defined in the paper by Shalev-Shwartz, Shammah and Shashua (2017). It seeks to standardize safety assurance in ways that are scalable for CAVs. RSS is a "white box" mathematical model for assuring the safety of driving. Their approach is to define mathematically five essential rules for safe driving:

1. Do not hit someone from behind.
2. Do not cut-in recklessly.
3. Right-of-way is given, not taken.
4. Be careful of areas with limited visibility.
5. If you can avoid an accident without causing another one, you must do it.

What does it tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

The mathematical approach to establishing safe driving behaviours is highly useful in the development of safe automated driving behaviours. The data computed in ensuring adherence to the essential rules could therefore form the basis of metrics for operational CAV safety.

Identified gap: However, the specific data that should be collected is not defined nor is there are any reference as to whether the five rules should be applied any differently depending on ODD.

7.9 Aptiv, Audi, Baidu, Continental et al. (2019), *Safety first for automated driving*

What approach is used in relation to evaluating (or specifying) CAV safe driving behaviour?

The report summarizes widely known safety by design and verification and validation (V&V) methods of SAE L3 and L4 automated driving to provide evidence of a positive risk balance for automated driving compared to average human driving performance. It is based on the input of OEMs, tiered suppliers and key technology providers, systematically breaking down safety principles into safety by design capabilities, elements and architectures and then to summarize the V&V methods in order to demonstrate the positive risk balance. With Level 3 and 4 automated driving systems still under development, the report represents guidance for potential methods and considerations in the development and V&V. The authors intended for the report to contribute to current activities working towards the industrywide standardization of automated driving.

The report lists a number of fail-safe and fail-degraded capabilities, which highlight information seen as necessary for safe automated driving. These include:

- determining current location;
- perceiving relevant objects;
- predicting the future behaviour of relevant objects;
- creating a collision-free and lawful plan;
- correct execution of the plan; and
- communicating with other road users.

It adopts the Sense-Plan-Act paradigm from mobile robotics in structuring automated driving performance.

What does it tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

The report is useful in citing ODD as an important element in automated driving operational safety. The fail-safe / fail-degraded capabilities highlight important elements that should be captured in assessing the capabilities of a CAV in driving safely.

Identified gap: The report does not discuss the data metrics that should be used to assess in-service safety of CAVs.

7.10 CertiCAV (2020-21), Connected Places Catapult and WMG, University of Warwick

What approach is used in relation to evaluating (or specifying) CAV safe driving behaviour?

CertiCAV was a UK Department of Transport funded project which was delivered by Connected Places Catapult and WMG, University of Warwick, UK. The project aimed to create a safety assurance framework for CAVs with a major focus on system safety. Key findings of the project included the creation of a behaviour library and subsequently a codified Highway Code for the UK. The project introduced the concept of “behavioural and tactical safety” as part of the safe driving behaviour of the CAV. Furthermore, the project defined nine principles that constitute good driving behaviour. These include avoiding collisions both as a contributing causal element (for the CAV) as well as when the CAV is not the cause of the accident. The project also proposed a linkage between the driving behaviour principles, codified Highway Code and the scenario generation process. However, the project does not include implications of cultural influences on the driving behaviour definition.

What does it tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

The project refers to behavioural and tactical safety concept as a function of the ODD. Furthermore, it suggests the codified Highway Code needs to be a function of the ODD with ODD attributes influencing not only the normal driving behaviour but also the selection and execution of the minimal risk manoeuvre.

Identified gap: The project does not include security requirements and requirements for remote operation. Safety and security metrics for both remain a gap for the industry. Furthermore, metrics for in-service monitoring were out of scope for the CertiCAV project.

7.11 Blumenthal, Fraade-Blanar, Best, Irwin (RAND Corporation) (2020), *Safe Enough*

What approach is used in relation to evaluating (or specifying) CAV safe driving behaviour?

The Safe Enough report by the RAND Corporation classified safety assessment into three aspects:

- Safety as a measurement: A data-driven approach to quantify safety performance
- Safety as a process: These are indicators of organization behaviour and include engineering processes involved in system-level verification and validation
- Safety as a threshold: Comparing “safety as a measurement” or “safety as a process” metrics to a given threshold.

Safety as a measurement: RAND Corporation in their previous report introduced the concept of "leading" (i.e. measures of pre-crash driving behaviour) and "lagging" (i.e. measures of crashes and post-crash outcomes) metrics. While lagging metrics may not be useful for initial approval of CAVs, the report suggests three types of indicators for the leading metrics: disengagements; legal infractions; and roadmanship. Roadmanship is a way to define good driving and can be seen akin to a codified Highway Code.

Safety as a process: In an attempt to gain trust and the fact that statistical comparisons may not be enough to demonstrate safety, *safety as a process* measures have been suggested to illustrate an organization's intent and commitment to making CAVs safe. To this end, safety cases are a way to demonstrate both rigour as well as well compliance with a process by documenting steps taken by the organization. Furthermore, compliance with standards has been suggested as a means to demonstrate "safety as a process".

Safety as a threshold: Defining a threshold (from measurement or process perspective) is an attempt to define "acceptably safe" as the minimum requirement that needs to be met or exceeded. Thresholds can be both quantitative (e.g., crash rates) and qualitative. Thresholds can be based on:

- Human Driving Performance (e.g. comparing to crashes in number of vehicle miles travelled);
- ADS Technology Potential (e.g. virtual driving test or subjective assessment of the AV technology); and
- Safety goals or objectives (e.g. ALARP, MEM, GAMAB).

What does it tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

Each of the aspects of safety assessment are deeply rooted in the ODD definition. From a "safety as a measurement" perspective, while ODD will not influence the measure's validity, it may influence how the measure is analysed and understood. From a "safety as a process" perspective, the level of rigour on the engineering processes needs to be proportional to the risk associated in the deployment ODD. From a "safety as a threshold" perspective, one would need to define new thresholds every time an ODD in expanded or modified.

Identified gap: While the report provides an elaborate suit of measures, it falls short in identifying the process of defining the measures from the perspective of an approval process.

7.12 World Economic Forum (2020), *Safe Drive Initiative/Autonomous Vehicle Policy Framework*

What approach is used in relation to evaluating (or specifying) CAV safe driving behaviour?

The World Economic Forum's Safe Drive Initiative aims to create a high-level framework to enable a regulator to perform an operational safety assessment with the ultimate goal to enabling the deployment of AVs without a safety driver. As part of this initiative, the Forum released four documents:

- *Creating safe autonomous vehicle policy*;
- *SafeDI scenario-based AV policy framework – an overview for policy makers*;
- *Taxonomy for Segmentation of Autonomous Delivery Vehicles and Personal Delivery Devices*; and
- *The Autonomous Vehicles Governance Ecosystem: A Guide for Decision-Makers*.

From a safe driving behaviour definition perspective, the framework document proposes a scenario-based approach and a concept of deployment ODD and milestone ODD. It proposes a four step (milestone) maturity process with milestone ODD being expanded gradually to ultimately be the same as the deployment ODD.

With each milestone, the fewer ODD parameters are restricted. For a scenario generation process, the framework suggests the combination of the ODD definition and a behaviour library be used to generate the qualitative scenarios (and subsequently use them for logical and concrete scenario definition). Furthermore, it suggests using knowledge-based and data-based approaches for scenario generation. In order to provide regulators and policy makers with a reference scenario set, the World Economic Forum has partnered with Safety Pool^{TM}, which has been founded by Deepen AI and WMG, University of Warwick. Safety Pool^{TM} Scenario Database provides scenarios for different ODDs and behaviour definitions.

What does it tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

The framework proposes the use of ODD taxonomy standards (e.g. PAS 1883) for the definition of the ODD which forms the first step in its safety assurance framework.

Identified gap: The framework does not provide mechanisms to judge completeness of scenario-based approaches. Furthermore, the framework does not include metrics or requirements on the security aspects of the CAV.

7.13 Wishart et al. (2020), *Driving safety performance assessment metrics for ADS-equipped vehicles*

What approach is used in relation to evaluating (or specifying) CAV safe driving behaviour?

This study summarizes a comprehensive literature review of metrics proposed for measuring the driving safety performance of both human-driven vehicles and automated vehicles, with the objective of developing a concise set of metrics that allow driving safety performance assessments to be effectively made and that align with the needs of both the ADS development and transportation engineering communities and accommodate differences in cultural/regional norms.

What does it tell us about the potential development of safety benchmarks of acceptable driving behaviour for CAVs in the UK across PAS 1883 ODDs?

The authors propose 14 metrics, each with justifications for inclusion:

- Driving Safety Performance Metrics.
  - Minimum Safe Distance Violation;
  - Proper Response Action;
  - Minimum Safe Distance Factor;
  - Minimum Safe Distance Calculation Error;
  - Collision Incident;
  - Rules of the Road Violation;
  - Achieved Behavioural Competency;
  - ADS Active;
  - Human Traffic Control Detection Error Rate; and
  - Human Traffic Control Violation Rate.
- Traffic Engineering-Developed Metrics.
  - Time-to-Collision;
  - Modified Time-to-Collision;
  - Post-Encroachment Time; and
  - Aggressive Driving.

They note the need for synchronized time stamping to ensure accurate analysis of data. However, aside from noting likely variations by jurisdiction and culture, they do not propose any adjustment to the metrics or thresholds based on ODD. In addition to objective measures of CAV safety, the authors highlight that the proposed metrics could be used to compare safety performance of a CAV relative to a human-driven vehicle.

Identified gap: The proposed metrics are highly relevant to the current study but do not account for ODD variations and the subjective/derived nature of the metrics may introduce challenges for CAV developers in producing the required data.

References

Aptiv, Audi, Baidu, BMW, Continental, Daimler, FCA, HERE, Infineon, Intel, and Volkswagen (2019). *Safety first for automated driving. White Paper*. Accessed April 2021 from: www.daimler.com/documents/innovation/other/safety-first-for-automated-driving.pdf

Balcombe, B. (2020). *Update on FG-AI4AD related activities.* Informal document GRVA-07-33, 7th GRVA, 21-25 September 2020, Agenda item 12(b). Accessed April 2021 from: https://unece.org/fileadmin/DAM/trans/doc/2020/wp29grva/GRVA-07-33e.pdf

Blumenthal, M. S., Fraade-Blanar, L., Best. R., Irwin, J. L. (2020). *Safe enough: approaches to assessing acceptable safety for automated vehicles.* RAND (RR-A569).

Bonnefon, J-F., Černý, D., Danaher, J., Devillier, N., Johansson, V., Kovacikova, T., Martens, M., Mladenovic, M.N., Palade, P., Reed, N., Santoni De Sio, F., Tsinorema, S., Wachter, S., Zawieska, K. (2020). *Ethics of Connected and Automated Vehicles Recommendations on road safety, privacy, fairness, explainability and responsibility. European Commission.*

California Department of Motor Vehicles (2020). *Article 3.7 – Testing of Autonomous Vehicles*. Accessed April 2021 from: www.dmv.ca.gov/portal/file/autonomous-vehicles-testing-without-a-driver-adopted-regulatory-text-pdf/

Connected Places Catapult (2021). *CertiCAV Software Framework Implementation and Evaluation*. Accessed May 2021 from: https://cp.catapult.org.uk/wp-content/uploads/2021/05/CertiCAV-software-framework-report-V1.0.pdf

Coupé, C., Oh, Y. M., Dediu, D., & Pellegrino, F. (2019). *Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche*. Science advances, 5(9), eaaw2594.

Crundall, D., Andrews, B., Van Loon, E., & Chapman, P. (2010). *Commentary training improves responsiveness to hazards in a driving simulator*. Accident Analysis & Prevention, 42(6), 2117-2124.

Department for Transport and Driver and Vehicle Standards Agency (2015). *The Highway Code*. London: The Stationery Office.

European Commission (2021). *Proposal for a Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts*. COM(2021) 206 final. Accessed May 2021 from: https://ec.europa.eu/newsroom/dae/redirection/document/75788

Finn, P., & Bragg, B. W. (1986). Perception of the risk of an accident by young and older drivers. *In: Accident Analysis & Prevention*, 18(4), 289-298.

Great Britain. Road Traffic Act 1988. London: The Stationery Office. Accessed March 2021 from www.legislation.gov.uk/ukpga/1988/52/contents

Harré, N., Foster, S., & O'Neill, M. (2005). Self-enhancement, crash-risk optimism and the impact of safety advertisements on young drivers. *In: British Journal of Psychology*, 96(2), 215-230.

Hatakka, M., Keskinen, E., Gregersen, N. P., Glad, A., & Hernetkoski, K. (2002). From control of the vehicle to personal self-control; broadening the perspectives to driver education. Transportation Research Part F. *In: Traffic Psychology and Behaviour*, 5(3), 201-215.

Helman, S., Grayson, G. B., & Parkes, A. M. (2010). *How can we produce safer new drivers? A review of the effects of experience, training and limiting exposure on the collision risk of new drivers*. TRL Insight Report INS005. Crowthorne: TRL Limited.

Ingenia (2015). Professor Paul Newman FREng. *Instilling robots with lifelong learning*. Accessed March 2021 from: www.ingenia.org.uk/Ingenia/Articles/37fa7f22-8b26-483c-b294-be7361de81e7

Isler, R. B., Starkey, N. J., & Williamson, A. R. (2009). Video-based road commentary training improves hazard perception of young drivers in a dual task. *In: Accident Analysis & Prevention*, 41(3), 445-452.

Marottoli, R. A., & Richardson, E. D. (1998). Confidence in, and self-rating of, driving ability among older drivers. *In: Accident Analysis & Prevention*, 30(3), 331-336.

McCormick, I. A., Walkey, F. H., & Green, D. E. (1986). Comparative perceptions of driver ability – A confirmation and expansion. *In: Accident Analysis & Prevention*, 18(3), 205-208.

McKenna, F. P., Horswill, M. S., & Alexander, J. L. (2006). Does anticipation training affect drivers' risk taking? *In: Journal of Experimental Psychology*. Applied, 12(1), 1.

McKenna, F. P., Stanier, R. A., & Lewis, C. (1991). Factors underlying illusory self-assessment of driving skill in males and females. *In: Accident Analysis & Prevention*, 23(1), 45-52.

MIIT (2021). *"Guidelines for the Management of Intelligent Connected Automobile Manufacturers and Product Access (Trial)" (Draft for Solicitation of Comments)*. First Division of Equipment Industry. Accessed May 2021 from: www.miit.gov.cn/jgsj/zbys/qcgy/art/2021/art_67412baef004441a9cafe0a440a928a2.html

Mills, K. L., Rolls, G. W. P., Hall, R. D., & McDonald, M. (1998). *The effects of hazard perception training. Department for Transport Research Report*. Accessed March 2021 from: https://eprints.soton.ac.uk/75643/

Nistér, D., Lee, H. L., Ng, J., & Wang, Y. (2019). *An introduction to the safety force field. NVIDIA White Paper*. Accessed April 2021 from: www.nvidia.com/content/dam/en-zz/Solutions/self-driving-cars/safety-force-field/an-introduction-to-the-safety-force-field-v2.pdf

Oxbotica (2017). *An interview with our founder, Prof Paul Newman*. Accessed March 2021 from: www.oxbotica.com/insight/an-interview-with-our-founder-prof-paul-newman/

The Police Foundation, Coyne, P. and Mares, P. (2020). *Roadcraft: The Police Driver's Handbook*. London: The Stationery Office.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *In: Nature Machine Intelligence*, 1(5), 206-215.

Royal Society for the Prevention of Accidents (2013). *Road safety and in-vehicle monitoring (black box) technology: Policy paper*. February 2013.

Shalev-Shwartz, S., Shammah, S. and Shashua, A. (2017). *On a formal model of safe and scalable self-driving cars*. Jerusalem: Mobileye.

United Nations Economic Commission for Europe (2020). *Proposal for a new UN Regulation on uniform provisions concerning the approval of vehicles with regards to Automated Lane Keeping System*. Geneva: UNECE. Accessed April 2021 from: https://undocs.org/ECE/TRANS/WP.29/2020/81

United Nations Economic Commission for Europe (2021). *Proposal for a new UN Regulation on Event Data Recorder*. ECE/TRANS/WP.29/2020/123/Rev.1. Geneva: UNECE. Accessed April 2021 from: https://unece.org/sites/default/files/2021-02/ECE-TRANS-WP29-2020-123r1e.pdf

Volvo (2019). *Will automation ever take off in construction?* Accessed March 2021 from: www.volvoce.com/global/en/news-and-events/news-and-stories/2019/will-automation-ever-take-off-in-construction/

Wishart, J., Como, S., Elli, M., Russo, B., Weast, J., Altekar, N. & James, E. (2020). *Driving Safety Performance Assessment Metrics for ADS-Equipped Vehicles.* SAE Technical Paper 2020-01-1206. Pennsylvania: SAE International.

389 Chiswick High Road
London W4 4AL
United Kingdom

Tel +44 (0)20 8996 9000
Fax +44 (0)20 8996 7400

info@bsigroup.com
www.bsigroup.com